

UNIVERSIDADE FEDERAL DO PARANÁ

LEANDRO FABIAN ALMEIDA ESCOBAR

**PÓS-PROCESSAMENTO DE PADRÕES PARA IDENTIFICAÇÃO DE
BENEFICIÁRIOS DE ALTO CUSTO EM OPERADORAS DE SAÚDE**

**CURITIBA
2015**

LEANDRO FABIAN ALMEIDA ESCOBAR

**PÓS-PROCESSAMENTO DE PADRÕES PARA IDENTIFICAÇÃO DE
BENEFICIÁRIOS DE ALTO CUSTO EM OPERADORAS DE SAÚDE**

Dissertação apresentada ao Programa de Pós-graduação em Ciência, Gestão e Tecnologia da Informação, Área de concentração: Gestão da Informação e do Conhecimento do Setor de Ciências Sociais Aplicadas da Universidade Federal do Paraná, como requisito parcial à obtenção do título de Mestre em Ciência, Gestão e Tecnologia da Informação.

Orientadora: Profª Drª Deborah Ribeiro Carvalho

**CURITIBA
2015**

Escobar, Leandro Fabian Almeida

Pós-processamento de padrões para identificação de beneficiários de alto custo em operadoras de saúde / Leandro Fabian Almeida Escobar. - 2015.

139 f.

Orientador: Deborah Ribeiro Carvalho.

Dissertação (Mestrado) – Universidade Federal do Paraná.

Programa de Pós- Graduação em Ciência, Gestão e Tecnologia da Informação, do Setor de Ciências Sociais Aplicadas.

Defesa: Curitiba, 2015

1. Gerenciamento da informação – Operadoras de saúde. 2. Mineração de dados (Computação). 3. Processo decisório. I. Carvalho, Deborah Ribeiro. II. Universidade Federal do Paraná. Setor de Ciências Sociais Aplicadas. Programa de Pós- Graduação em Ciência. Gestão e Tecnologia da Informação. III. Título.

CDD 658.4038

TERMO DE APROVAÇÃO

Leandro Fabian Almeida Escobar

**“PÓS- PROCESSAMENTO DE PADRÕES PARA IDENTIFICAÇÃO DE
BENEFICIÁRIOS DE ALTO CUSTO EM OPERADORAS DE SAÚDE”**

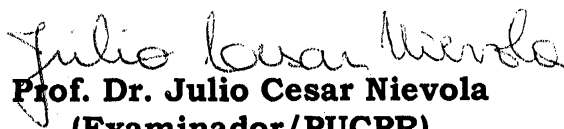
**DISSERTAÇÃO APROVADA COMO REQUISITO PARCIAL PARA
OBTENÇÃO DO GRAU DE MESTRE NO PROGRAMA DE PÓS-
GRADUAÇÃO EM CIÊNCIA, GESTÃO E TECNOLOGIA DA INFORMAÇÃO
DA UNIVERSIDADE FEDERAL DO PARANÁ, PELA SEGUINTE BANCA
EXAMINADORA:**



**Prof.^a Dr.^a Deborah Ribeiro Carvalho
(Orientadora/PUCPR)**



**Prof.^a Dr.^a Denise Fukumi Tsunoda
(Examinadora/UFPR)**



**Prof. Dr. Julio Cesar Nievola
(Examinador/PUCPR)**

26 de março de 2015

Dedico esse trabalho a Vimara, Isadora e Marcelo, por servirem como âncoras em minha vida e pelo apoio sempre disponível, em os quais eu não teria a força necessária para termina-lo.

AGRADECIMENTOS

À minha orientadora, Profa. Dra Deborah Ribeiro Carvalho, pelo acompanhamento, orientação, amizade e dedicação, sem o quais este trabalho não seria possível.

Ao programa Programa de Pós-graduação em Ciência, Gestão e Tecnologia da Informação da Universidade Federal do Paraná, na pessoa de sua coordenadora Profa Drz Denise Fuukumi Tsonoda. Também por suas contribuições e sugestões no trabalho.

Aos Professores Leandro Henrique Souza e Alexandre Denes, pelo apoio e estímulo em todos os momentos do trabalho.

Ao Prof Marcelo Rossano Dallagassa, pelas contribuições e sugestões quanto ao protótipo desenvolvido.

Ao Prof Julio Cesar Nievola, pelas contribuições e sugestões no trabalho.

RESUMO

As operadoras de saúde armazenam grandes volumes de dados sobre procedimentos realizados por seus beneficiários e, na mesma medida que criam valiosas oportunidades, também geram dificuldades em analisar e interpretar os padrões contidos em tais bases de dados. A Mineração de Dados pode constituir uma alternativa para compensar tal deficiência, mas sua adoção e uso cotidiano no suporte à decisão na área da Saúde ainda são baixos. O problema de pesquisa abordado se relaciona ao baixo uso do processo de descoberta de conhecimento (KDD) na área e questiona-se até que ponto é possível fomentar sua adoção na rotina dos especialistas em Saúde. Assim, este trabalho apresenta um modelo que fomente o uso do KDD na rotina da gestão em Saúde, mediante a identificação de beneficiários de alto custo em operadoras de Saúde, avaliando sua aceitação junto a especialistas na área. Foram identificados os pontos de atenção recomendados na literatura para uso efetivo dos resultados do KDD na Saúde, o modelo foi elaborado e um protótipo foi desenvolvido com a capacidade de obter padrões e pós processá-los, obtendo padrões de sequência de janelas temporais. Sua aceitação foi avaliada junto aos especialistas, tendo atingido 78% de aceitação. Destaca-se que todos os especialistas envolvidos concordam que o modelo é aplicável à sua rotina de trabalho. Conclui-se que a associação dos pontos de atenção para o uso da Mineração de Dados na Saúde e a atenção a requisitos específicos dos especialistas, mediante o desenvolvimento de um modelo que facilite a obtenção e a exploração em padrões minerados, se mostrou capaz de fomentar o uso do KDD na rotina em Saúde.

Palavras chave: Descoberta de conhecimento; Mineração de Dados, Apoio à decisão em Saúde.

ABSTRACT

Health insurance companies store large amounts of data on procedures performed over its beneficiaries and to the same extent that create valuable opportunities also lead to difficulties in analyzing and interpreting the standards contained in such databases. Data mining can be an alternative to compensate for this deficiency, but their adoption and use in everyday decision support in healthcare are still low. The research problem addressed is related to the low use of knowledge Discovery in data bases process (KDD) in the area and wonders to what extent it is possible to foster its adoption in experts routine. So, this researche presents a model that promotes the use of KDD in the routine management in Health, by identifying high-cost beneficiaries and evaluating its acceptance by the experts. Points of attention for the use of Data Mining in Health were identified in the literature for effective use of the results of KDD. A model and a prototype were developed with the capability of discovering patterns and post process into sequential and time frames patterns. Its acceptance was evaluated with the experts, reaching 78% of acceptance. It is noteworthy that all the experts involved agree that the model is applicable to their daily work. We conclude that the association of points of attention for the use of Data Mining in Health and attention to specific requirements of experts, by developing a model to facilitate the obtaining and exploring mined patterns, has been shown to promote the use of KDD in routine Health.

Keywords: Knowledge discovery; Data Mining, Decision Support in Health.

LISTA DE FIGURAS

Figura 1 Visão geral do KDD.....	24
Figura 2 Exemplo de Classificação	27
Figura 3 Exemplo de três grupos identificados em uma base de dados	28
Figura 4 Interdependência das etapas do método proposto	36
Figura 5 Etapas e procedimentos da pesquisa	37
Figura 6 Representação do modelo proposto	46
Figura 7 Arquitetura geral do modelo proposto	46
Figura 8 Fragmento de arquivo com procedimentos de interesse, a ser importado pelo modelo.....	59
Figura 9 Sequência de arquivos gerados e processamento para obtenção dos padrões sobre procedimentos de alto custo.....	61
Figura 10 Fragmentos dos arquivos de entrada para Apriori e AssocTemp gerados pelo modelo.....	61
Figura 11 Fragmento de arquivo gerado pelo Apriori durante a criação da base de conhecimento pelo modelo.....	62
Figura 12 Fragmento de arquivo gerado pelo Chrono_Assoc durante a criação da base de conhecimento pelo modelo	62
Figura 13 Exemplo de arquivos com padrões de intervalo de tempo gerados para um procedimento de interesse pelo modelo.....	63
Figura 14 Fragmento de arquivo gerado pelo AssocTemp	63
Figura 15 Modelo de entidades e relacionamentos do banco de dados do modelo proposto	64
Figura 16 Classes implementadas no modelo.....	65
Figura 17 Caixa de diálogo Carregar arquivo com procedimentos.....	67
Figura 18 Interface para geração da base de conhecimento do modelo.....	68
Figura 19 Diagrama de atividades - analisar padrões de procedimentos.....	70
Figura 20 Interface de usuário do modelo	71
Figura 21 Ficha médica do beneficiário.....	71
Figura 24 Sequência de navegação no modelo proposto	77

LISTA DE QUADROS

Quadro 1 Resumo dos tipos de decisão em saúde	22
Quadro 2 Frequência absoluta e relativa dos procedimentos segundo a categorização de custo entre janeiro de 2006 e janeiro de 2014	39
Quadro 3 Frequência absoluta dos procedimentos de Interesse indicados por especialista.....	39
Quadro 4 Padrões obtidos para o procedimento 40301109AMB - ÁCIDO LÁTICO..	40
Quadro 5 Grupo de requisitos funcionais para modelo	41
Quadro 6 Frequência dos pontos de atenção associados com Auxílio ao Planejamento em Saúde	41
Quadro 7 Procedimentos de interesse indicado pelo especialista	49
Quadro 8 Construtos e respectivas questões para a avaliação do modelo proposto	50
Quadro 9 Chaves de respostas para as questões relativas ao perfil dos entrevistados	51
Quadro 10 Pontos de atenção associados com Representação visual dos resultados.	54
Quadro 11 Pontos de atenção associados com Descrição de eventos ocorridos.	54
Quadro 12 Pontos de atenção associados com Desenvolvimento de funcionalidade específica.	55
Quadro 13 Pontos de atenção associados com Fornecimento de subsídios para explicações causais.	56
Quadro 14 Requisitos e justificativas apontadas pelos especialistas em apresentação preliminar.....	56
Quadro 15 Layout do arquivo de entrada com os procedimentos liberados pela operadora de saúde	59

LISTA DE GRÁFICOS

Gráfico 1	Frequência dos pontos de atenção na bibliografia analisada	35
Gráfico 2	Distribuição dos especialistas quanto à formação profissional	75
Gráfico 3	Pontuação obtida nas questões da avaliação do modelo proposto	76
Gráfico 4	Pontuação obtida pelos construtos avaliados para o modelo proposto.	78
Gráfico 5	Pontuação do construto "Capacidade de descrever os eventos" em função da área de formação dos especialistas.	79
Gráfico 6	Pontuação do construto "Apresentação de padrões e informações" em função da área de formação dos especialistas.	80
Gráfico 7	Pontuação do construto "Satisfação do especialista quanto ao modelo" em função da área de formação dos especialistas.	81

LISTA DE TABELAS

Tabela 1 Exemplos de regras de associação sequenciais e não sequenciais obtidas pelo modelo.....	72
Tabela 2 Padrões descobertos para a regra de associação: DOSAGEM DE TRANSAMINASE OXALACETICA e GASOMETRIA -> ÁCIDO LÁTICO	73
Tabela 3 Padrões descobertos para a regra de associação: CULTURA AUTOMATIZADA e DOSAGEM DE FOSFATASE ALCALINA -> ÁCIDO LÁTICO ...	73
Tabela 4 Padrões descobertos para a regra de associação: MARCADORES TUMORAIS e DOSAGEM DE FOSFATASE ALCALINA -> ANTÍGENO CARCINOEMBRIOGÊNICO.....	74
Tabela 5 Padrões descobertos para a regra de associação: DOSAGEM DE UREIA e ZOFRAN -> ANTÍGENO CARCINOEMBRIOGÊNICO”	74

LISTA DE SIGLAS

DRE	Descobre Regras de Exceção
KDD	<i>Knowledge Discovery in Databases</i>
PIBIC	Programa Institucional de Bolsas de Iniciação Científica
PUCPR	Pontifícia Universidade Católica do Paraná
SAD	Sistema de Apoio à Decisão
SI	Sistema de Informação
UFPR	Universidade Federal do Paraná

SUMÁRIO

1 .INTRODUÇÃO	16
1.1 Objetivos.....	18
1.1.1 Objetivo geral.....	18
1.1.2 Objetivos específicos	19
2 .REFERENCIAL TEÓRICO	20
2.1 Sistemas de apoio à decisão	20
2.2 Decisão na saúde.....	21
2.3 Descoberta de conhecimento em bases de dados	23
2.4 Tarefas da mineração de dados	26
2.4.1 Classificação.....	26
2.4.2 Agrupamento	27
2.4.3 Regras de associação	28
2.5 Pós-processamento	29
2.6 Pontos de atenção para a adoção da descoberta de conhecimento em bases de dados na saúde	32
3 .MÉTODO	36
3.1 Levantamento dos requisitos específicos dos especialistas	37
3.2 Construção e prototipação do modelo	40
3.2.1 Arquitetura do modelo proposto.....	45
3.3 Preparação da base de conhecimento para avaliação do modelo	48
3.4 Avaliação do modelo proposto.....	49
3.5 Aspectos éticos	52
4 .APRESENTAÇÃO E DISCUSSÃO DOS RESULTADOS	53
4.1 Pontos de atenção para adoção do kdd de dados na saúde	53
4.2 Requisitos dos especialistas	56

4.3 Modelo proposto	58
4.3.1 Importação de dados e geração da base de conhecimento	58
4.3.2 Interação do usuário com o modelo.....	67
4.4 Padrões obtidos e base de conhecimento	72
4.4.1 Interpretação dos padrões obtidos	72
4.5 Avaliação do modelo proposto	75
4.5.1 Pontuação obtida pelas questões do instrumento de avaliação	76
4.5.2 Pontuação obtida pelos construtos do instrumento de avaliação	78
5 . CONSIDERAÇÕES FINAIS	82
REFERÊNCIAS	85
APÊNDICES	89

1 INTRODUÇÃO

O uso da Tecnologia da Informação favorece o crescimento do volume de dados armazenados, exigindo assim, maneiras eficazes e eficientes para a extração de informações de tais bases de dados, de forma a apoiar as decisões quanto ao diagnóstico, à prevenção de doenças ou à aplicação eficiente de recursos. No entanto a utilização desses dados por gestores e especialistas ainda é limitada, comprometendo as chances de extração de novos conhecimentos, dada a dificuldade em analisar grandes volumes de dados (SANTANA *et al.* 2006).

Na área da Saúde, bem como nas demais áreas, a decorrente dificuldade em interpretar e avaliar esse volume de dados demanda a utilização de processos que automatizem a extração eficiente e eficaz de informações oportunas, permitindo análises mais complexas para a tomada de decisão (TRINDADE, 2005; KOBUS, 2006; SANTANA *et al.* 2006; LOPES, 2007; DALLAGASSA, 2009).

Desta forma, uma alternativa para compensar a dificuldade em analisar os dados disponíveis na Saúde é prover os especialistas de ferramentas que obtenham modelos para apoio à descrição de eventos, explorando as bases de dados e os sistemas de informação disponíveis, combinando a capacidade estatística e probabilística do aprendizado de máquina com o conhecimento dos especialistas (TRINDADE, 2005).

Assim, o processo de descoberta de conhecimento em bases de dados (KDD – *Knowledge Discovery in Databases*) (FAYYAD, PIATETSKY-SHAPIRO e SMYTH, 1996), tem demonstrado seu potencial em facilitar e aumentar a produção de conhecimento em vários domínios de negócios e isto também pode ocorrer na área da Saúde (SHILLABEER, 2005).

A descoberta de padrões com o KDD pode levar a identificação de informações relevantes em bases de dados, fornecendo modelos para o planejamento de ações preventivas ou para a tomada de decisão quanto à população em questão (RAMÓN, 2007; DALAGASSA, 2009; MEYFROIDT *et al.*, 2009), aumentando a eficácia de tratamentos e políticas de prevenção bem como a eficiência na aplicação de recursos em tais iniciativas (DALAGASSA, 2009).

Apesar de todo desenvolvimento já realizado, sua aplicação continua sendo objeto de pesquisas por novas soluções que se aproximem do real interesse dos

potenciais usuários (MARISCAL, *et al*, 2010) e que estabeleçam critérios específicos da área da Saúde para sua adoção rotineira.

Mariscal *et al* (2010) argumentam que as razões pela constante pesquisa se devem ao fato de que, obstante a existência de estudos científicos quanto ao uso de KDD, sua adoção nos processos de tomada de decisão diários não é ampliada por conta da pouca familiaridade dos especialistas com a metodologia e suas vantagens e dificuldades (MEYFROIDT *et al* 2009), a prevalência de estudos estatísticos com objetivo de revelar relações lineares entre os fatores de saúde (CRUZ-RAMIREZ *et al* 2012) ou, ainda, a exploração comparativa de técnicas de Mineração de Dados, deixando de lado etapas relativas à interpretação e avaliação de resultados (BLONBERG, 2010) e sua consequente contribuição para o dia-a-dia do especialista em Saúde.

O problema em questão configura-se no fato de que, embora as pesquisas sejam constantes, os resultados do KDD não são adotados extensivamente na rotina da gestão em Saúde e que, para fomentar sua adoção cotidiana, é necessário identificar quais fatores devem ser atendidos de forma que os resultados do KDD constituam instrumentos para a tomada de decisão na área.

É neste contexto que emerge a questão central desta pesquisa: É possível fomentar o uso do KDD na rotina da gestão da Saúde?

As respostas a esta questão giram em torno da análise dos elementos que facilitam ou dificultam a adoção efetiva dos padrões obtidos com o KDD para apoio à tomada de decisão na área da Saúde.

Esta pesquisa se apoia, então, no pressuposto de que a baixa adoção do KDD na decisão em Saúde limita a capacidade de obtenção de conhecimentos por meio dos padrões presentes nas bases de dados disponíveis.

Pressupõe-se, também, que a adoção do KDD cumpre com a função primária da informação em Saúde, que é detectar problemas individuais e coletivos do quadro sanitário de uma população, oferecer elementos que subsidiem a análise rigorosa desse quadro e, então, apresentar alternativas para minimizar a situação encontrada. Para isto, em termos ideais, abrange tanto informações relativas ao binômio “saúde e doença” como as de natureza administrativa, em acordo com Targino (2009).

Enfim, o alinhamento dos resultados do KDD com as necessidades dos especialistas da área da Saúde pode oferecer subsídios para o ganho de efetividade

das decisões, beneficiando diretamente gestores, equipes especialistas, beneficiários e pacientes.

Por outro lado, a pesquisa esta pesquisa está limitada pela aplicação em uma operadora de Saúde, ensejando que sua aplicação em mais operadoras pode esclarecer mais requisitos, melhorar técnicas e trazer novas perspectivas sobre o ponto de atenção para a mineração de dados na Saúde. Outra limitação é aquela relativa ao estudo baseado na gestão em saúde. Estudos relativos à decisão clínica podem enriquecer as conclusões, representando novos desafios e ainda, ampliar a contribuição científica e social da pesquisa aqui apresentada.

A contribuição científica deste projeto está na organização dos pontos de atenção para a adoção do processo de descoberta de conhecimento em bases de dados na rotina em Saúde, bem como a proposta de um modelo que permita a navegação em padrões obtidos com KDD, de forma a facilitar a identificação de beneficiários de alto custo, servindo como base para encaminhamentos metodológicos, de processos e de uso de técnicas válidos para o desenvolvimento de sistemas para apoio à decisão na Saúde. A organização do processo de avaliação, seu critérios, forma de aplicação e limites para aceitação também contribuem para a realização de pesquisas futuras, uma vez que traçam um caminho, ainda que tênue, para a avaliação de sistemas de apoio à decisão baseados na mineração de dados.

Socialmente, o projeto contribui com a proposta de um modelo que melhore a apropriação das bases de dados para o apoio à decisão na Área da Saúde, mediante a mineração de dados e pós-processamento de padrões de maneira eficiente e possibilitando o aumento da capacidade em elaborar planos de prevenção e de tratamento de doenças mediante modelos de apoio à decisão que contribuam para o ganho de eficiência de organizações de Saúde e para ganhos na da qualidade de vida da população por conta da melhoria das condições sanitárias individuais ou coletivas.

1.1 Objetivos

1.1.1 Objetivo geral

Propor um modelo que fomente o uso do KDD na rotina da gestão em Saúde, mediante a identificação de beneficiários de alto custo em operadoras de Saúde.

1.1.2 Objetivos específicos

- a) levantar os pontos de atenção recomendados na literatura para uso efetivo dos resultados do KDD na Saúde;
- b) prototipar um modelo que permita a utilização de padrões obtidos com o KDD para apoio à decisão em Saúde;
- c) avaliar o modelo junto aos especialistas de uma operadora de saúde do Paraná.

1.1.3 Organização do trabalho

Além da introdução, o segundo capítulo deste trabalho apresenta os conceitos, pressupostos, técnicas e proposições que servem como sustentação do método da pesquisa. São abordados os princípios sobre sistemas de apoio à decisão, mineração de dados, processo de descoberta de conhecimento em bases de dados e pós-processamento de padrões.

O terceiro capítulo descreve o método da pesquisa, percorrendo sua organização em etapas mediante a definição das técnicas, estruturas, atividades, modelos e respectivas justificativas para realização dos objetivos de pesquisa e busca às respostas para a questão de pesquisa.

O quarto capítulo discorre sobre os resultados alcançados, descrevendo tanto a implementação do modelo proposto na forma de protótipo quanto o processo de avaliação, analisando os detalhes sobre as considerações dos especialistas envolvidos.

O quinto capítulo encerra o trabalho, sem a pretensão de encerrar o assunto, dissertando sobre as percepções, resultados e trabalhos futuros ensejados pela pesquisa.

2 REFERENCIAL TEÓRICO

Este capítulo apresenta os conceitos relativos aos Sistemas de Apoio à Decisão, ao KDD, à Mineração de Dados e suas aplicações. Também relata a aplicação do KDD para apoio à decisão na área da Saúde, de forma a explorar os aspectos técnicos da Mineração de Dados em Saúde relevantes para a pesquisa atual.

2.1 Sistemas de apoio à decisão

O conceito tradicional de decisão está relacionado com a opção pela alternativa mais satisfatória entre várias e o estabelecimento de uma série de ações para a realização de tal alternativa, tratando, assim, da resolução de um problema (SHIMIZU, 2010).

A riqueza dos cenários e a quantidade de informações disponíveis requer o uso de recursos computacionais, denominados sistemas de apoio à decisão (SAD), cujo propósito é apoiar o processo de tomada de decisão (SPRAGUE e WATSON, 1991; TURBAN e ARONSON 2001; TURBAN *et al*, 2009) e que permitem o acesso interativo aos dados, que ajudam os responsáveis pela tomada de decisão a utilizar dados e modelos para análise adequada e, então, formularem decisões mais eficientes e eficazes (TURBAN *et al*, 2009).

Sprague e Watson (1991) destacam que algumas das características de um SAD são:

- ser voltado para solução de problemas menos estruturados e menos especificados com os quais os trabalhadores se deparam;
- combinar o uso de modelos ou técnicas analíticas e funções tradicionais de acesso e recuperação de informações;
- ser de fácil uso por pessoal não especializado em computação, enfatizando sua interatividade;
- ser adaptável e flexível, acomodando mudanças no ambiente e na abordagem do usuário para a tomada de decisão.

Com a difusão do armazenamento eletrônico de dados sobre pacientes e beneficiários e seus respectivos históricos de saúde, sistemas de apoio à decisão em Saúde podem se tornar valiosos recursos no aumento da acurácia do diagnóstico, da qualidade e da eficiência nos tratamentos (MANGIAMELI *et al*, 2004;

KOBUS, 2006; DALLAGASSA, 2009). O uso da Tecnologia da Informação na área da Saúde melhorou significativamente a prática clínica. Em particular, o uso de prontuários eletrônicos e sistemas de apoio à decisão contribuem para o aumento da habilidade dos especialistas em localizar, recuperar e filtrar as informações necessárias para um desempenho mais efetivo, apoiando na integração de informações relevantes e no descarte de informações irrelevantes e, desta forma, suportando o processo de decisão (PATEL *et al*, 2013).

Ainda que o especialista seja competente, sem o apoio de tecnologia, não é possível integrar e assimilar todas as informações necessárias para a tomada de decisão. O uso de Sistemas de apoio à decisão auxilia na solução dos desafios enfrentados pelos especialistas frente aos volumes de informações que devem ser manipuladas para a efetiva tomada de decisão (PATEL, V. *et al*, 2013).

2.2 Decisão na saúde

A tomada de decisão na área da Saúde, por sua vez, é tanto um resultado quanto um componente das atividades desenvolvidas e compreender o que influencia positivamente no processo de tomada de decisão torna-se imperativo para a consistência de eventuais propostas de sistemas de apoio à decisão.

A decisão em saúde pode ser dividida em Decisão Clínica Curativa, Decisão Preventiva e Decisão para Aplicação de Recursos (MANGIAMELI *et al*, 2004; KOBUS, 2006; SANTANA *et al* (2006); DALLAGASSA, 2009, NÉRI *et al*, 2011).

A decisão clínica curativa está relacionada com diagnóstico, a determinação do tratamento a ser habilitado e com a segurança do paciente durante tal tratamento (NÉRI *et al*, 2011). Trata-se da resposta necessária uma vez que a doença está instalada e identificada. Para tanto, o médico coleta dados do paciente de maneira a formular um diagnóstico e definir um conjunto de medidas curativas (MANGIAMELI *et al*, 2004; KOBUS, 2006).

Assim, a decisão clínica curativa gira em torno da determinação precisa de que o paciente está doente e da aplicação do tratamento eficaz para tal mal (MANGIAMELI *et al*, 2004).

A decisão preventiva está relacionada com a medicina preventiva e trata do mapeamento da situação sanitária, indicando a prevalência de doenças e de riscos à saúde atual e futura de indivíduos ou de uma determinada população, servindo

como base para a elaboração de estratégias de prevenção e promoção da saúde (KOBUS, 2006; DALLAGASSA, 2009).

A decisão para aplicação de recursos trata do problema de escassez de recursos, uma vez que o agravamento das doenças relativas ao trabalho e ao envelhecimento da população requer que a aplicação de recursos seja mais eficiente, exigindo que as organizações de Saúde conheçam profundamente os ciclos de atendimento, prevenção de doenças, diagnósticos e gerenciamento de casos como bases para o aumento da qualidade no atendimento e a redução dos custos (MANGIAMELI *et al*, 2004; KOBUS, 2006; SANTANA *et al*, 2006; DALLAGASSA, 2009).

Os tipos de decisão identificados para a Saúde e seus aspectos gerais são mostrados no Quadro 1.

Tipos de decisão	Aspectos gerais
Decisão clínica curativa	Trata do diagnóstico e da determinação do tratamento que melhor atenda às necessidades específicas do paciente
Decisão preventiva	Trata da determinação dos riscos aos qual uma população está exposta e a determinação de estratégias para promoção da saúde.
Aplicação de recursos	Trata da eficiência na gestão da saúde, mediante estratégias que permitam a otimização do uso de recursos, como a antecipação de casos complexos, a auditoria de procedimentos e a criação de estratégias de prevenção.

Quadro 1 Resumo dos tipos de decisão em saúde

Fonte: Mangiameli et al (2004); Kobus (2006); Santana et al (2006), Dallagassa (2009) e Néri et al (2011)

Smith *et al* (2008) advertem que a tomada de decisão em saúde é um processo de maior complexidade, uma vez que os especialistas enfrentam contextos dinâmicos, utilizam um corpo de conhecimento crescente e com múltiplas variáveis e indivíduos envolvidos. Ao contrário de possibilidades isoladas, as decisões em Saúde são constituídas por ciclos decisão-ação onde as situações evoluem e durante o qual decisões e ações influenciam-se mutuamente.

Patel *et al* (2013), por sua vez, resumem o processo de tomada de decisão em saúde como uma combinação de indução e dedução, chamada raciocínio abduutivo, no qual o especialista gera explicações plausíveis de acordo com os dados observados e testa tais explicações, na busca de comprovar suas hipóteses.

Smith *et al* (2008) ainda consideram que a incerteza permeia todo o processo de decisão em Saúde, uma vez que o contexto é caracterizado pelo fato de que nem toda a informação necessária para a decisão está disponível ou é conhecida.

Compreender como a tomada de decisão em Saúde se dá, quais são suas características e quais são os domínios nos quais os diferentes tipos de decisão em Saúde estão inseridos permite determinar quais estratégias podem ser adotadas para a elaboração de sistemas para apoio à decisão na área.

Embora esta pesquisa esteja voltada para o fomento do uso do KDD na rotina em Saúde, o que permitiria sua aplicação em qualquer dos tipos de decisão, o modelo proposto é aplicado à identificação de beneficiário de alto custo, contribuindo para a eficiência tanto do processo de identificação quanto para as ações decorrentes disto. Assim, o foco da aplicação do modelo proposto está voltado para a decisão para aplicação de recursos.

2.3 Descoberta de conhecimento em bases de dados

A descoberta de conhecimento em bases de dados (KDD) consiste no processo interativo, iterativo e não trivial de descoberta de informações, mediante a identificação de padrões válidos, novos, potencialmente úteis e compreensíveis em dados (FAYYAD *et al*, 1996). Este processo inclui as etapas expressas na Figura 1.

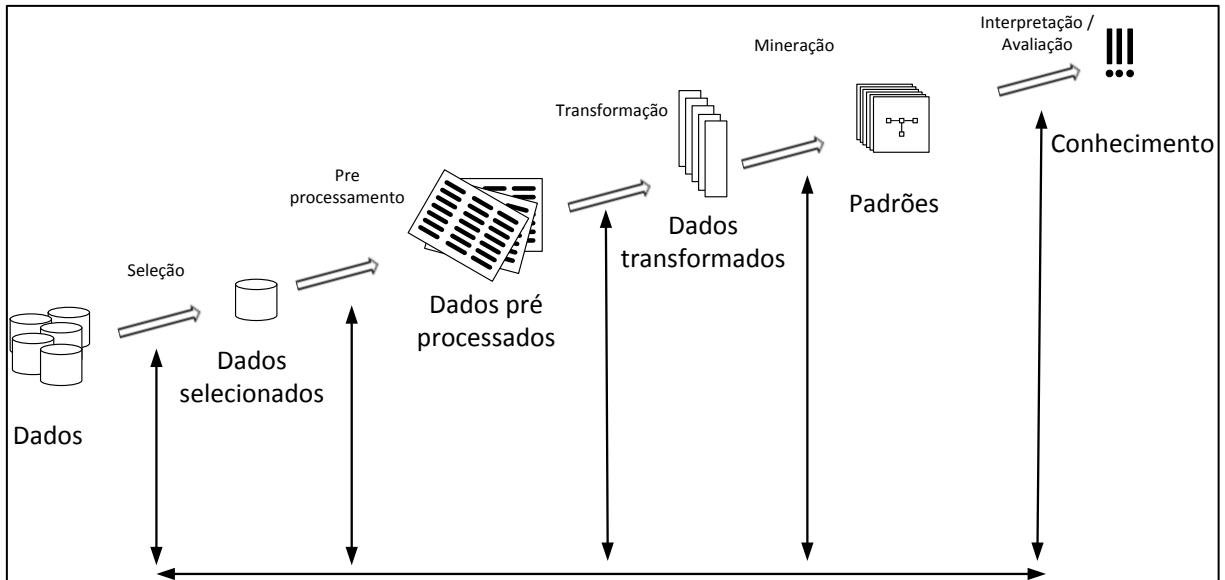


Figura 1 Visão geral do KDD
 Fonte: Fayyad *et al* (1996)

- **seleção:** criação de um conjunto inicial de atributos visando à descoberta de conhecimentos com Mineração de Dados;
- **pré-processamento:** envolve a remoção de ruídos, preenchimento de dados nulos (não disponíveis) e tratamento de eventuais inconsistências;
- **transformação:** trata da geração dos dados para uso dos mesmos pelos algoritmos de Mineração de Dados;
- **mineração:** tarefa de descoberta de padrões mediante algoritmos de Mineração de Dados;
- **interpretação/avaliação:** Análise e entendimento dos padrões obtidos frente ao cenário do qual se originaram, estabelecendo como tais descobertas contribuem para as decisões decorrentes.

A seleção tem como finalidade selecionar os dados que representam o domínio do problema em questão e a partir dos quais a Mineração de Dados será realizada. Nesta etapa, cria-se um conjunto de registros que representem os fatos em estudo, armazenados nas fontes de dados disponíveis.

O pré-processamento visa realizar a limpeza dos ruídos, preenchimento de dados em branco, remoção de inconsistências, tratamento de dados ausentes, seja

pelo preenchimento ou pela remoção de dados ou tuplas, e a organização dos dados por conta de relações cronológicas (temporalidade).

Han e Kamber (2006) destacam as seguintes estratégias para pré-processamento dos dados:

- **limpeza**: visa eliminar as eventuais inconsistências encontradas nos dados, de modo que não influenciem nos resultados dos algoritmos utilizados para a Mineração.;
- **integração**: é comum que os dados a serem minerados sejam oriundos de diversas fontes e torna-se necessário integrá-los de forma a elaborar um repositório único e consistente;
- **redução**: o volume dos dados utilizados na mineração pode ser alto e inviabilizar análises e o próprio desempenho dos algoritmos. Nestes casos, a massa de dados pode ser convertida em um conjunto menor, mas sem perder a representatividade original.

A transformação trata da conversão dos dados de forma a atender a tarefa específica de mineração dos dados, reduzindo dimensões e facilitando a extração de características dos fatos representados.

Alguns algoritmos trabalham com valores categorizados, outros somente com valores numéricos. Nestes casos, é necessário transformar valores originalmente numéricos em categorias ou valores nominais em contínuos, dependendo do caso. Han e Kamber (2006) recomendam o uso de técnicas de agrupamento e generalização para a criação de categorias e técnicas de normalização ou criação de novos atributos para a geração de valores contínuos.

A Mineração de Dados, por sua vez, consiste na aplicação de algoritmos de análise e descoberta que, dentro de limitações computacionais, produzem um conjunto específico de padrões sobre os dados, também denominados modelos. (FAYYAD, PIATETSKY-SHAPIO e SMYTH, 1996).

Constituída de vários métodos, a Mineração de Dados realiza tarefas que buscam descrever os padrões, como associações existentes entre os dados, ou classificar uma determinada instância, ou amostra, dos dados em uma classe previamente identificada mediante modelos previamente desenvolvidos a partir dos padrões presentes nos dados. (CARVALHO, 2005; HAN e KAMBER, 2006; BELAZZI e ZUPAN, 2008; MILOVIC e MILOVIC, 2012).

A interpretação dos resultados da Mineração de Dados, por sua vez, garante que conhecimentos úteis sejam derivados tanto dos dados quanto dos padrões encontrados neles (FAYYAD *et al*, 1996).

Mitra e Acharya (2003) colocam que a interpretação dos resultados deve ser apoiada por recursos de visualização e navegação, a fim de aumentar sua inteligibilidade, colaborando com a identificação de padrões realmente interessantes para o usuário. Neste processo de navegação, as variáveis podem ser filtradas e simulações devem ser permitidas, de modo que o especialista possa experimentar não só os modelos, mas os resultados obtidos (MILOVIC e MILOVIC, 2012).

O KKD é, então, a base da aplicação e do funcionamento do modelo proposto. Esse processo orienta as etapas do modelo de forma a contribuir para a eficiência e eficácia na identificação de beneficiários de alto custo.

2.4 Tarefas da mineração de dados

A Mineração de Dados é a etapa do KDD responsável por realizar uma ou mais das seguintes tarefas: Classificação, Agrupamento, Descoberta de Regras de Associação (MITRA E ACHARYA. 2003; GENG e HAMILTON, 2006).

2.4.1 Classificação

Tem por objetivo classificar os registros (fatos ou eventos) em uma classe predeterminada, mediante a obtenção de modelos que preveem um objetivo (considerado atributo meta) a partir dos dados. A classificação é um processo de obtenção de um modelo que determina a qual classe, dentre várias, uma nova instância (novo exemplo de registros) pertence (HAN e KAMBER, 2006; MILOVIC e MILOVIC, 2012).

Cada classe corresponde a padrões de comportamento dos itens da base de dados (atributos previsores). Formalmente, seja T o conjunto de todos os exemplos selecionados e C o conjunto de todas as classes identificadas em T . A cada classe C_i , corresponde uma descrição D_i , dos atributos selecionados. Assim, mediante essas descrições, é possível elaborar um modelo que descreva um exemplo e do conjunto T , como pertencente à classe C_i , quando tal exemplo satisfaz a descrição D_i (CARVALHO, 2005).

Os modelos classificadores são construídos particionando recursivamente o espaço de dados, dividindo-o em áreas mais específicas, de forma a separar as classes, conforme a Figura 2 (CARVALHO, 2005).

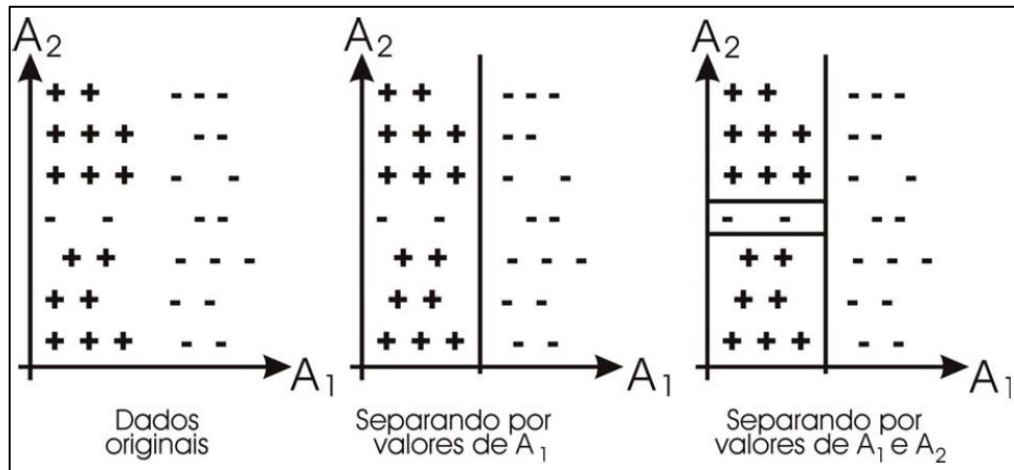


Figura 2 Exemplo de Classificação

Fonte: Freitas e Lavington (1998) *apud* Carvalho (2005)

2.4.2 Agrupamento

A tarefa de agrupamento mapeia os registros em um, de vários conjuntos que representem conjuntos de registros baseados na sua similaridade, segmentando-se os dados heterogêneos em grupos homogêneos (MITRA e ACHARYA 2003). O resultado é um conjunto finito de grupos que ocupam o mesmo espaço de classificação (CHEN *et al*, 1996; CARVALHO, 2005).

Isso permite que as características comuns sejam sumarizadas e utilizadas para formar a descrição dos dados (CARVALHO, 2005). Diferentemente da classificação, esta técnica não requer que os registros sejam previamente classificados e pode ser importante tanto na redução dos dados quando na identificação da similaridade entre as diferentes instâncias (HAN e KAMBER, 2006). Carvalho (2005), inclusive, destaca a possibilidade de que o agrupamento seja realizado durante a análise inicial dos dados, para posterior tarefa de classificação (Figura 3).

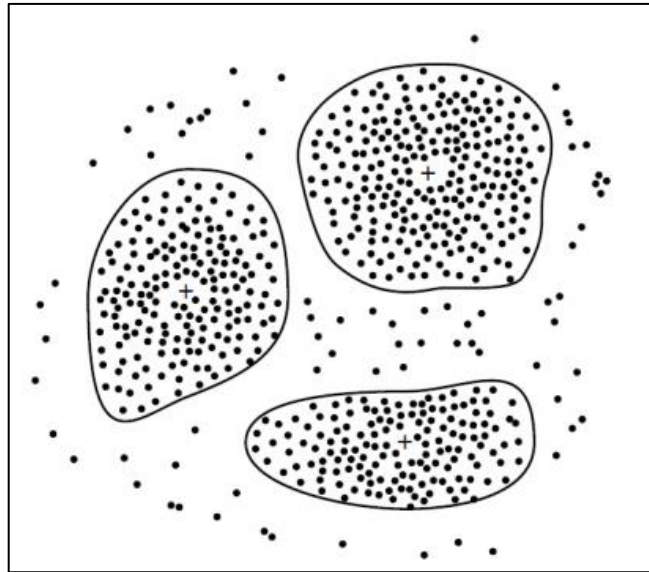


Figura 3 Exemplo de três grupos identificados em uma base de dados
 Fonte: Han e Kamber (2006)

2.4.3 Regras de associação

Trata da descoberta de relações de associação ou correlações frequentes entre os diferentes atributos da base de dados e são representadas na forma de regras do tipo Se X então Y ($X \rightarrow Y$). Sendo X e Y representantes dos itens de dados do conjunto de entrada em que $X \cap Y = \emptyset$. X é o antecedente da regra e Y é o conseqüente da regra. As regras de associação descrevem o padrão de ocorrência de X e Y aparecendo juntos na mesma transação (AGRAWAL e SRIKANT, 1994, CARVALHO, 2005, BORGELT, 2010).

Para cada regra de associação descoberta são determinados o suporte e a confiança da regra.

O suporte (Expressão 1) representa a porcentagem de transações na base de dados que contém os itens da regra. Desta forma, o suporte indica a quantidade de transações que contém os itens da regra. A confiança da regra (Expressão 2) representa a quantidade de registros que possuem antecedente X e também possuem o conseqüente Y. (AGRAWAL e SRIKANT, 1994).

O suporte representado pelo percentual das transações que contém os itens da regra na base de dados $X \cup Y$ em relação ao total de transações T , e pode ser entendido como a probabilidade de que todos os itens da regra estejam presentes em uma transação:

$$suporte = \frac{X \cup Y}{T} \quad (1)$$

A confiança (Expressão 2) é representada pela proporção de transações que contém os itens da regra em relação à proporção de transações que possuem o antecedente da regra (BORGELT, 2010) e pode ser entendida como a probabilidade de se encontrar o consequente Y uma vez que o antecedente X for encontrado:

$$confiança = \frac{suporte(X \cup Y)}{suporte(X)} \quad (2)$$

Em contrapartida, Borgelt (2010) implementa em sua versão do Apriori um suporte diferente daquele implementado por Agrawal e Srikant (1994). O autor implementa o suporte como sendo a proporção de transações que possuem o antecedente da regra (Expressão 3):

$$suporte = \frac{X}{T} \text{ (BORGELT, 2010)} \quad (3)$$

A implementação do modelo da pesquisa aqui apresentada utiliza o conceito de suporte preconizado por Borgelt (2010), demonstrando a quantidade (relativa ou absoluta) de transações que possuem o antecedente da regra, indicado a quantidade de transações às quais a regra pode ser aplicada.

No modelo proposto, a interpretação das regras deve observar suporte e confiança em conjunto, uma vez que um suporte e uma confiança altos representam associações mais comuns e possivelmente conhecidas pelo especialista. Por outro lado, medidas baixas para suporte e altas para a confiança podem revelar procedimentos pouco frequentes, porém cujas associações sejam mais fortes, ou seja, menos evidentes para o especialista com uma probabilidade maior de que antecedente e consequente ocorram em conjunto.

2.5 Pós-processamento

O pós-processamento contempla a depuração e/ou síntese dos padrões descobertos (CARVALHO *et al*, 2012) após a Mineração dos Dados, ou ainda, a descoberta de relações específicas dentro dos padrões obtidos, que podem

evidenciar aqueles de maior interesse (HUSSAIN *et al*, 2000), reduzindo o volume a ser avaliado pelo especialista.

Ainda que os padrões obtidos estejam diretamente ligados aos fatos, o contexto no qual estão inseridos determina seu interesse (SILBERSCHATZ e TUZHILIN, 1996; GENG e HAMILTON, 2006), ensejando, por exemplo, questões relativas à descoberta de exceções quanto aos padrões mais frequentes ou às sequências de causa e efeito e ao padrão de distribuição dos eventos no decorrer do tempo.

Neste sentido, Hussain *et al* (2000), propõem o pós-processamento de regras de associação mediante a obtenção de regras de exceção, organizando o conjunto de regras obtidas em regras gerais (com suporte e confiança altos) e regras de exceção, cujo conseqüente nega a regra geral e suporte é baixo. Carvalho *et al* (2012) aplicam a descoberta de regras de exceção em uma base de dados de pacientes de fisioterapia e encontram padrões interessantes no tratamento de encurtamento de ligamentos.

Milani e Carvalho (2013) propõem um método para extração de regras de exceção, tendo desenvolvido o DRE (Descobre Regras de Exceção), que recebe as regras de associação obtidas como Apriori (BORGELT, 2010) e identifica as regras de exceção conforme Hussain *et al* (2000).

Já Roddick e Spiliopoulou (1999) discute a ordenação de eventos em cadeias causais, identificando padrões em que X leva a Y, propondo, desta forma, o termo mineração temporal, em que são descobertos padrões ordenados na linha do tempo.

Gomes e Carvalho (2011) propõem um método híbrido para a extração de regras sequenciais, tendo desenvolvido o algoritmo Chrono_Assoc, que recebe as regras de associação geradas pelo algoritmo Apriori (BORGELT, 2010) e determina, então, as medidas de suporte e confiança cronológicas.

Segundo Gomes e Carvalho (2011) o suporte cronológico trata a frequência de ocorrências do antecedente e do conseqüente nesta ordem I_s , em relação ao número total de transações na base T (Expressão 4). A confiança cronológica representa a frequência dos itens da regra tomados em sequência I_s e em relação ao conjunto total de regras que possuem tais itens I (Expressão 5).

$$SuporteCronológico(t) = \frac{I_s}{T} \quad (4)$$

$$ConfiançaCronológica(t) = \frac{I_s}{I} \quad (5)$$

Desta forma, uma regra é considerada sequencial quando seu suporte cronológico e sua confiança cronológica são maiores que 0 (zero).

Sokoloski, Carvalho e Dallagassa (2014) propõem uma estratégia, desenvolvendo o algoritmo Assoctemp, para pós-processar regras de associação avaliando a sequência e os intervalos de tempo relativos à ocorrência de eventos.

O Assoctemp recebe as regras de associação obtidas com o Apriori (BORGELT, 2010) e um arquivo de transações em que os itens ou eventos estão acompanhados do momento em que ocorrem, seja a data ou a hora. O intervalo de tempo sobre o qual se deseja pós-processar os padrões – janela temporal de interesse – é informado pelo usuário. O resultado do processamento são medidas de suporte e confiança temporais, demonstrando a frequência dos eventos dentro da janela temporal de interesse (SOKOLOSKI, CARVALHO e DALLAGASSA, 2014).

O suporte temporal₁ (Expressão 6) demonstra a proporção de transações que contem o antecedente e o consequente, respectivamente, I_s em relação ao número de transações na base T . Esta métrica retrata a frequência de transações às quais o padrão temporal se aplica.

$$SuporteTemporal_1 = \frac{I_s}{T} \quad (6)$$

A confiança temporal₁ (Expressão 7) representa a proporção de transações que possuem o antecedente e o consequente dentro da janela temporal de interesse I_j , respectivamente, em relação ao número transações que contém o antecedente e o consequente, respectivamente, I_s .

$$ConfiançaTemporal_1 = \frac{I_j}{I_s} \quad (7)$$

O suporte temporal₂ (Expressão 8) representa a razão entre o número de aparições de sequências – antecedente e consequente, respectivamente I_s e o total de transações na base T . Esta métrica retrata a densidade do padrão temporal em

relação às transações, dado que uma transação pode conter a mesma sequência de eventos mais de uma vez.

$$SuporteTemporal_2 = \frac{Is}{T} \quad (8)$$

A confiança temporal₂ (Expressão 9) representa a razão entre o número de aparições de sequências dentro da janela temporal de interesse Isj e o número de aparições de sequências Is .

$$ConfiançaTemporal_2 = \frac{Isj}{Is} \quad (9)$$

O pós-processamento permite, assim, facilitar a avaliação e obter outros padrões a partir dos padrões originalmente obtidos, aumentando as chances de identificar aqueles que venham a ser úteis e agreguem conhecimento para o especialista. (CARVALHO *et al*, 2012).

2.6 Pontos de atenção para a adoção da descoberta de conhecimento em bases de dados na saúde

Entender os pontos de atenção para o uso dos resultados do KDD na Saúde permite a elaboração de táticas para aumento da eficiência e maior aproveitamento dos benefícios gerados, dando maiores condições de que os padrões obtidos com a Mineração de Dados e o pós-processamento de padrões sejam utilizados na rotina dos especialistas, mediante a combinação de recomendações e aplicações já experimentadas e avaliadas nos estudos científicos disponíveis.

Os pontos de atenção foram adotados conforme Carvalho, Escobar e Tsunoda (2014), fruto de estudo e publicação dentro do escopo desta pesquisa, para os quais se apresenta uma síntese do conteúdo integral, disposto no Apêndice A.

Analisando as 18 publicações selecionadas a partir dos critérios de inclusão, percebe-se que 75% relatam experimentos envolvendo a área clínica e apenas 25% envolvendo gestão da saúde.

Sobre as três etapas do KDD, todos os autores relatam Mineração de Dados, 14 autores também discutem a etapa de pré-processamento (78%) e 7 autores discutem o pós-processamento (39%),

A relação de pontos de atenção obtida a partir da leitura e análise das 18 publicações é:

- **Apropriação de protocolo específico:** atenção aos processos específicos de diagnóstico ou tratamento de acordo com indicação formal da especialidade médica a partir de protocolos ou diretrizes amplamente aceitas pela comunidade especializada;
- **Avaliação da qualidade dos dados:** trata da atenção e crítica à consistência, disponibilidade e tratamento dos dados. Este ponto de atenção está diretamente relacionado às etapas de seleção e preparação dos dados. Ainda que a qualidade dos dados faça do problema do desenvolvimento de Sistemas de Informação em geral, sua presença se justifica pela necessidade de se melhorar a qualidade da coleta de dados, mesmo ocorrendo em sistemas específicos ou ocasião diferente do momento da tomada de decisão em Saúde;
- **Integração de diferentes bases de dados:** Enriquecimento da base de dados a partir de diferentes fontes, sejam elas distribuídas, em formatos diversos ou ainda pertencentes às organizações diferentes da organização objeto do estudo. Um exemplo é a necessidade de complementar dados sobre atendimentos com dados sobre o perfil demográfico das regiões amostras do estudo;
- **Desenvolvimento de funcionalidade específica:** Construção de software específico para uso ou exploração dos modelos gerados. Desenvolvimento de sistema ou interface de usuário que suporte o processo de tomada de decisão tanto para a visualização dos dados quanto para a navegação nos padrões encontrados;
- **Obediência a processo de trabalho específico:** O sistema de coleta de dados e de apoio à decisão reproduz ou segue um conjunto de procedimentos encadeados logicamente e definidos como modo de trabalho dentro da organização em saúde. Trata-se da reprodução de processo de trabalho particular;
- **Utilização de modelo de informações específicas:** Adoção de conjunto de informações específicas e previamente definidas seja por protocolos ou pela própria organização Saúde. Trata-se da determinação de quais informações são relevantes sem que sejam

indicadas as sequências lógicas dos procedimentos ou ainda a ordem das atividades de atendimento;

- **Avaliação (subjetiva ou objetiva) da relevância dos padrões encontrados:** Avaliar os padrões extraídos determinando o quanto são relevantes ou interessantes para o especialista em saúde.
- **Combinação de diferentes tipos de tarefas de Mineração de Dados:** Uso de mais de uma tarefa de Mineração de Dados para a realização do experimento ou identificação de padrões;
- **Deteção de padrões em tempo real:** Extrair e apresentar padrões simultaneamente à ocorrência dos eventos que são fontes dos dados.
- **Auxílio ao planejamento em saúde:** Apoiar ações com base nos modelos extraídos pela Mineração de Dados;
- **Representação visual dos resultados:** Criação de recursos que facilitem a compreensão dos modelos extraídos mediante a visualização dos resultados. Trata da necessidade de comunicação amigável dos padrões encontrados para os especialistas em Saúde envolvidos nos estudos;
- **Descrição de eventos ocorridos numa determinada população:** Descrever padrões aos quais uma determinada amostra está sujeita, explicando as relações encontradas entre os dados minerados.
- **Previsão da ocorrência de eventos numa determinada população:** Antecipar eventos de acordo com os padrões identificados nas bases de dados;
- **Fornecimento de subsídios para explicações causais:** Demonstração das cadeias de eventos, suas causas e consequências.

A frequência em que tais pontos de atenção são citados nos relatos (Gráfico 1), demonstrando que a Previsão da Ocorrência de Eventos, com 11 citações, é mais frequente que a Descrição de Eventos Ocorridos, com 6 citações.

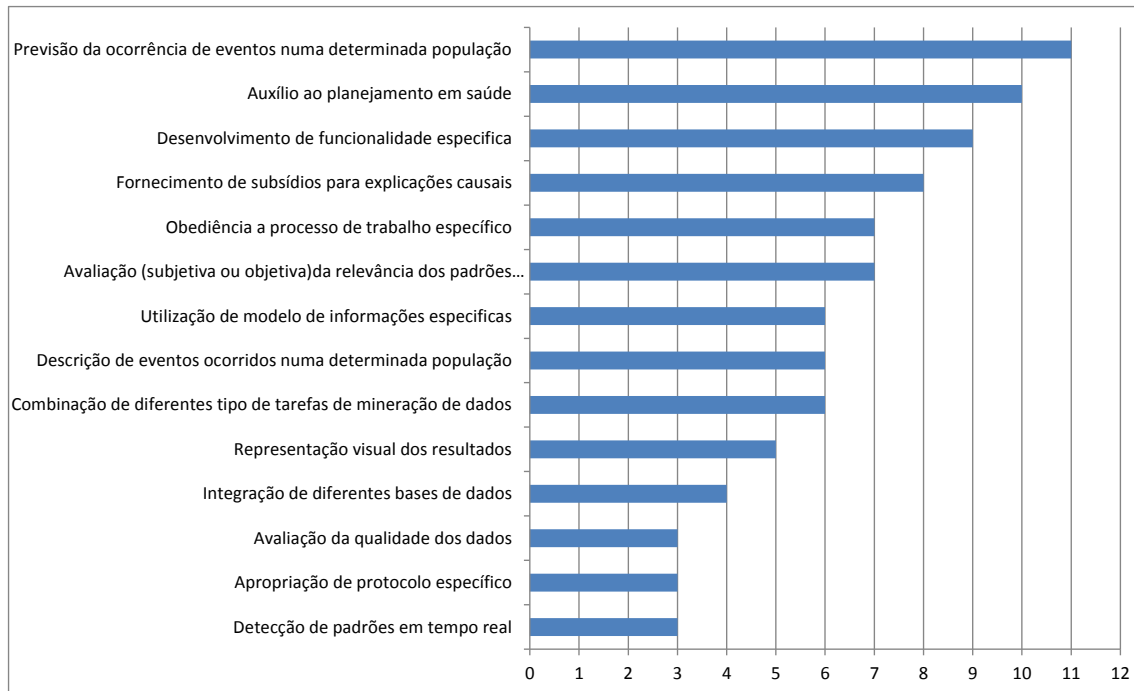


Gráfico 1 Frequência dos pontos de atenção na bibliografia analisada
Fonte: Carvalho, Escobar e Tsunoda (2014).

Com a finalidade de evidenciar quais pontos de atenção estão mais fortemente relacionados entre si, os pontos de atenção foram submetidos ao algoritmo Apriori (AGRAWAL e SRIKANT, 1994) e ao algoritmo DRE – Descubra Regras de Exceção, desenvolvido por Milani e Carvalho (2013), conforme Hussain et al (2000).

Uma filtragem baseada nos Pontos de Atenção demonstrou quais associações são relevantes para o planejamento da aplicação do KDD, quais pontos podem ser combinados para ganho de efetividade e uso na rotina de decisão em Saúde.

3 MÉTODO

Esta pesquisa relaciona-se com o uso cotidiano do KDD na gestão em Saúde e apoia-se na identificação de beneficiários de alto custo em operadoras de saúde, como aplicação prática que permita a avaliação do modelo proposto e sua capacidade de fomentar o uso do KDD na rotina em saúde.

Assim, o estudo em questão envolve uma pesquisa qualitativa, descritiva, aplicada e de raciocínio indutivo de forma a identificar os fatores que contribuem para a adoção sistemática do KDD na Saúde, bem como propor um modelo baseado na descoberta de conhecimento em bases de dados, que permita a navegação sobre estes padrões descobertos.

Buscando a uniformização de termos, gestores, técnicos, especialistas ou quaisquer outras pessoas que possam ou venham a contribuir com o entendimento das especificidades do domínio Saúde, serão denominados “especialistas da área da Saúde” ou tão somente “especialistas”, generalizando os papéis atuantes na área da Saúde, de forma a facilitar o entendimento quando se tratar de declarar qual papel está envolvido no contexto.

A pesquisa aqui apresentada está organizada em três etapas, partindo dos pontos de atenção recomendados na literatura selecionada e específica ao KDD aplicado na Saúde, passando pela prototipação e finalizando na avaliação do modelo proposto. Tais etapas estão conectadas criando uma relação de interdependência baseada nos seus resultados preliminares (Figura 4).

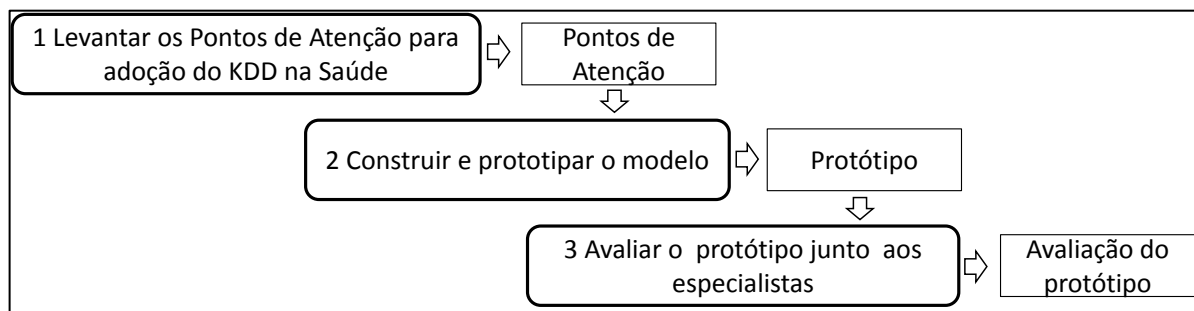


Figura 4 Interdependência das etapas do método proposto
Fonte: O autor (2014)

Por conta desta interdependência, surgem resultados individuais que, para fins de organização do documento, são demonstrados na descrição de cada etapa, nesta seção.

Cada etapa, por sua vez, está organizada em uma série de procedimentos, os quais constituem o caminho para sua realização e efetivação de seus resultados (Figura 5).

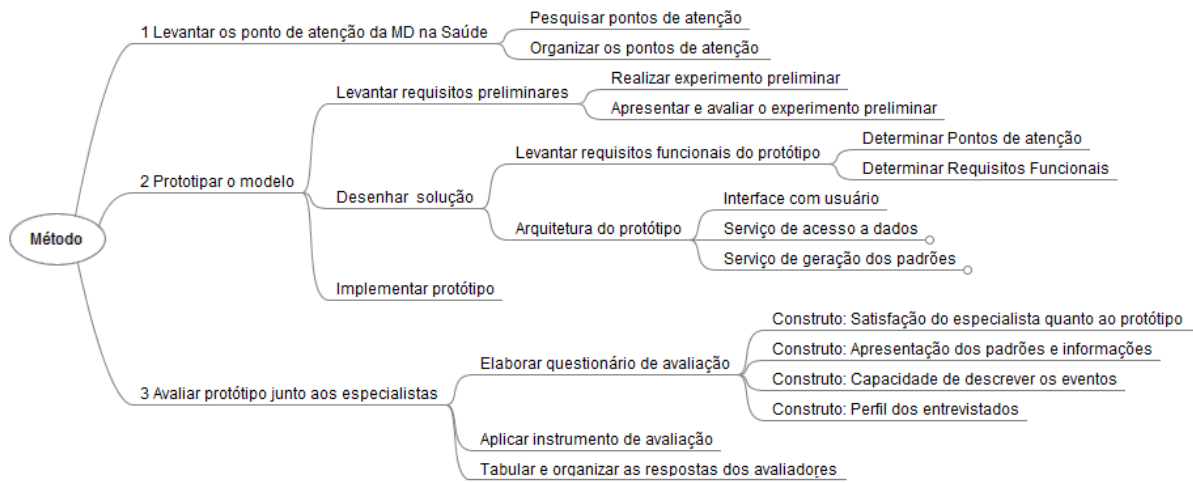


Figura 5 Etapas e procedimentos da pesquisa
Fonte: O autor (2014)

A Etapa 1 - Levantar Pontos de Atenção para adoção do KDD na Saúde trata-se de uma pesquisa sistemática com o objetivo de identificar os pontos de atenção recomendados pela literatura que facilitem o uso do KDD na rotina da Saúde. Tais pontos de atenção servem como subsídio para prototipação do modelo proposto, de forma que este seja mais aderente às necessidades dos especialistas e, desta forma, tenha mais chances de uso rotineiro.

A Etapa 2 – Construir e prototipar o modelo, trata-se da identificação dos requisitos dos especialistas que compõem o contexto da aplicação e validação do modelo e, ainda, da elaboração do desenho e construção do modelo proposto, endereçando os pontos de atenção (oriundos da Etapa 1) e os requisitos dos especialistas na forma de lógica a ser implementada.

A Etapa 3 – Avaliar o modelo junto aos especialistas: trata da avaliação do modelo proposto, de forma a verificar a aplicabilidade do modelo na rotina dos especialistas envolvidos.

3.1 Levantamento dos requisitos específicos dos especialistas

Um experimento preliminar foi elaborado com o objetivo de facilitar a apresentação para os especialistas, buscando estabelecer uma base para a

discussão sobre padrões minerados e, desta forma, obter uma visão inicial dos requisitos funcionais para o modelo a ser desenvolvido.

O experimento em questão foi orientado à gestão em saúde, a partir de bases de dados oriundas das liberações de procedimentos relativos a beneficiários de uma operadora de saúde suplementar paranaense e teve o objetivo responder aos especialistas qual é o padrão de realização dos eventos (realização de procedimentos) relativos aos beneficiários e que significam tratamentos de alto custo, permitindo a elaboração de políticas que facilitem o diagnóstico de doenças, reduzindo o número de procedimentos necessários e, desta, forma, aumentando a eficiência dos recursos à disposição dos beneficiários.

Aplicou-se o processo KDD (FAYYAD, PIATETSKY-SHAPIRO e SMYTH, 1996) com a intenção de obter regras de associação com o algoritmo Apriori (AGRAWAL e SRIKANT, 1994) e pós-processando de maneira a obter padrões sequenciais, com o uso do Chrono_Assoc (GOMES e CARVALHO, 2011) e dos padrões sequenciais dentro de intervalos de tempo com o do Assoctemp (SOKOLOSKI, CARVALHO e DALLAGASSA, 2014) no intuito de obter a descrição dos padrões de ocorrência de procedimentos em beneficiários de alto custo.

A base de dados utilizada nesta etapa contem 55.722 liberações de procedimentos realizados entre janeiro de 2006 e janeiro de 2014 relativas a 248 beneficiários. Os registros possuem informações sobre o gênero, idade e o histórico de procedimentos realizados pelos beneficiários, bem como a sua identificação de acordo com a categorização prévia de custos, apresentada pela operadora de saúde, sendo: Altíssimo Custo, Casos Complexos, Doente Crônico, Diagnosticado e Bem Estar (Quadro 2). As liberações de procedimentos que não estavam categorizadas quanto ao custo, totalizando 1378 liberações, foram definidas com o valor “N/D”, a fim de mantê-las no experimento.

Foram obtidas 13.649 regras de associação, pós-processadas, em conjunto com os dados originais, mediante o algoritmo Chrono_Assoc (GOMES e CARVALHO, 2011) e com o AssocTemp (SOKOLOSKI, CARVALHO e DALLAGASSA, 2014), para a obtenção dos padrões de associação cronológica dos procedimentos selecionados, buscando obter possíveis relações de causa e efeito na sua realização.

Classificação de custo	Total	Freq. %
1-Altissimo_Custo	18.302	33%
2-Casos_Complexos	12.889	23%
3-Doente_Cronico	10.877	20%
4-Diagnosticado	6.093	11%
5-Bem_Estar	6.183	11%
ND	1.378	2%
TOTAL	55.722	

Quadro 2 Frequência absoluta e relativa dos procedimentos segundo a categorização de custo entre janeiro de 2006 e janeiro de 2014

Fonte: O autor (2015)

Para viabilizar a análise de regras com potencial interesse, o especialista indicou oito procedimentos de interesse (Quadro 3), que representam tratamentos de alto custo, servindo como filtros para as análises dos padrões obtidos.

Procedimentos		Total
Código	Descrição	
10101039AMB	CONSULTA EM PRONTO SOCORRO	2207
40301109AMB	ACIDO LACTICO (LACTATO)	60
40302016AMB	ANTIGENO CARCINOEMBRIOGÊNICO (CEA), DOSAGEM	164
40304507AMB	GASOMETRIA (PH, PCO2, SA, O2, EXCESSO BASE)	10
40316122AMB	HERCEPTIN - 440mg	453
40316378AMB	MARCADORES TUMORAIS (CA 19.9, CA 125, CA 72-4, CA 15-3, ETC.) CADA, DOSAGEM	611
40805018AMB	PROTEINA C, DOSAGEM	362
92403034MED	TORAX - 1 INCIDENCIA	237

Quadro 3 Frequência absoluta dos procedimentos de Interesse indicados por especialista
Fonte: O autor (2014)

Os padrões obtidos com o programa Apriori (BORGELT, 2010) e pós-processados com o Chrono_Assoc (GOMES e CARVALHO, 2011) e AssocTemp (SOKOLOSKI, CARVALHO e DALLGASSA, 2014) foram consolidados com o auxílio de planilhas eletrônicas e filtrados de acordo com os procedimentos de interesse.

Os padrões de associação, padrões cronológicos e janelas de tempo para o procedimento Ácido Láctico são demonstrados no Quadro 4. Todos os demais procedimentos de interesse foram avaliados levando em conta a mesma estrutura.

Consequente	Precedente	Sup Apriori	Conf Apriori	Sup Chrono	Conf Chrono	Conf	Conf	Conf	Conf	Conf	Conf
						10 dias	30 dias	60 dias	120 dias	240 dias	365 dias
40301109AMB	20104049AMB	1,10%	66,70%	0,37%	33,33%	0,00%	0,00%	0,00%	0,00%	100,00%	100,00%
40301109AMB	90844351MED	1,50%	50,00%	0,73%	50,00%	0,00%	50,00%	100,00%	100,00%	100,00%	100,00%
40301109AMB	79944957MAT	1,50%	50,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
40301109AMB	79979025MAT	1,50%	50,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
40301109AMB	92030073HOS	1,50%	50,00%	0,73%	50,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
40301109AMB	50010212AMB	1,50%	75,00%	1,10%	75,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
40301109AMB	40301567AMB	1,50%	50,00%	0,73%	50,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
40301109AMB	71065296MAT	1,50%	50,00%	0,73%	50,00%	0,00%	0,00%	0,00%	0,00%	50,00%	50,00%

Quadro 4 Padrões obtidos para o procedimento 40301109AMB - ÁCIDO LÁTICO

Fonte: O autor (2014)

Os requisitos e necessidades apontados pelos especialistas foram registrados de forma a serem incorporados ao modelo (Quadro 5).

3.2 Construção e prototipação do modelo

O modelo proposto nesta pesquisa tem a finalidade de sistematizar os pontos de atenção para adoção da Mineração de Dados e os requisitos dos especialistas, permitindo a navegação nos padrões obtidos com o processo de descoberta de conhecimento em bases de dados – KDD - de forma a facilitar a identificação de beneficiários indicados para auditoria de procedimentos de alto custo, de forma que tais padrões sejam utilizados na rotina dos especialistas.

Os requisitos do modelo proposto foram organizados em dois grupos. O primeiro grupo é formado pelos pontos de atenção para uso da Mineração de Dados considerados aplicáveis ao modelo. O segundo grupo contém os requisitos funcionais desdobrados das necessidades apontadas junto aos especialistas durante a apresentação do experimento preliminar, de maneira que o modelo possa ser utilizado pelos especialistas ou auditores.

O Quadro 5 contém os requisitos funcionais para o modelo, definidos pelos especialistas, agrupados e descritos para um melhor entendimento do modelo proposto.

Os pontos de atenção endereçados para o modelo foram selecionados tendo em vista que a identificação dos padrões de eventos de alto custo está relacionada com a necessidade de apoiar decisões relativas ao planejamento de estratégias para promoção da qualidade de vida ou para a redução de custos em tratamentos.

Assim, as regras de associação foram filtradas de acordo com o ponto de atenção “Auxílio ao planejamento em saúde”, totalizando 57 regras e foram contados os pontos de atenção que estavam associados (Quadro 6).

Grupo	Requisitos
Grupo 1: Pontos de Atenção	PA01 - Desenvolvimento de funcionalidade específica
	PA02 - Representação visual dos resultados
	PA03 - Descrição de eventos ocorridos
	PA04 - Fornecimento de subsídios para explicações causais
Grupo 2 – Requisitos funcionais	RF01 - Carregar dados desnormalizados com os registros de eventos oriundos das operações da empresa.
	RF02 – Mapear os atributos que representam os beneficiários, procedimentos e as datas de ocorrência.
	RF03 - Gerar lista classificada de acordo com a frequência dos eventos da base de dados.
	RF04 - Obter padrões de associação, cronologia e janelas de tempo.
	RF05 - Navegar nos padrões obtidos <ul style="list-style-type: none"> - filtrar regras de associação que possuem o evento selecionado no antecedente, no consequente ou ambos; - exibir todas as regras ou somente implicações; - exibir os eventos coocorrentes com o evento selecionado.
	RF06 - Exibir os dados relativos aos beneficiários, filtrados de acordo com os padrões selecionados.

Quadro 5 Grupo de requisitos funcionais para modelo
Fonte: O autor (2014)

Ponto de Atenção	Frequência absoluta	Frequência relativa
Descrição de eventos ocorridos	41	72%
Avaliação da qualidade dos dados	24	42%
Fornecimento de subsídios para explicações causais	20	35%
Avaliação (subjetiva ou objetiva) da relevância dos padrões encontrados	11	19%
Representação visual dos resultados	7	12%

Quadro 6 Frequência dos pontos de atenção associados com Auxílio ao Planejamento em Saúde
Fonte: O autor (2014)

O ponto “Avaliação da qualidade dos dados” foi desconsiderado porque o modelo proposto exige a entrada dos dados já preparados para a extração de padrões, não requerendo que os dados sejam analisados quanto à sua qualidade.

O ponto de atenção “Avaliação (subjettiva ou objetiva) da relevância dos padrões encontrados” também foi desconsiderado por estar além do escopo desta pesquisa, devendo ser abordado em trabalhos futuros para continuidade da pesquisa no que tange a organização e avaliação dos demais pontos de atenção para o uso do KDD.

Desta forma, os pontos de atenção adotados para o modelo proposto ficaram organizados da seguinte maneira:

PA01 - Desenvolvimento de funcionalidade específica: Mediante a prototipação de uma ferramenta que abstraia o processo de Mineração de Dados e que represente os padrões obtidos, permitindo a navegação nos padrões e suas respectivas análises.

PA02 – Representação visual dos resultados: A representação dos padrões deve ser tal que permita uma leitura fácil e que o especialista consiga perceber as relações entre os diferentes eventos sem maiores necessidades de treinamento. Ou seja, o modelo deve ser amigável.

PA03 - Descrição de eventos ocorridos: A obtenção de regras de associação associadas com o pós-processamento de padrões cronológicos e de janelas de tempo permite descrever os padrões de ocorrência dos eventos (procedimentos) relativos às transações na operadora de saúde.

PA04 - Fornecimento de subsídios para explicações causais: Os padrões cronológicos e as janelas de tempo permitem a dedução da influência entre eventos. Levanto o especialista a identificar quais eventos são causas e seus respectivos efeitos.

Os requisitos funcionais, organizados no Grupo 2 são assim especificados:

RF01 - Carregar dados desnormalizados com os registros de eventos oriundos das operações da empresa: O modelo deve ser capaz de ler um arquivo

do tipo texto, separado por ponto e vírgula, no qual a primeira linha deve conter os nomes dos atributos e as demais linhas devem conter os dados relativos à ocorrência dos eventos a serem minerados.

RF02 – Mapear os atributos que representam os beneficiários, procedimentos e as datas de ocorrência: O modelo deve extrair todos os atributos lidos do arquivo de carga e armazená-los na base de dados de maneira a garantir sua utilização tanto na obtenção dos padrões, servindo com entrada para os algoritmos de mineração, quanto na navegação do especialista. Tais atributos representam:

- a) eventos: Eventos são ocorrências relativas às operações no sistema de informações operacionais da empresa e sobre os quais o especialista tem interesse em obter padrões mediante Mineração de Dados.
- b) data de ocorrência: Tratam-se das datas em que os eventos ocorrerem e que devem estar devidamente representadas no arquivo de carga de dados.
- c) beneficiário: Trata-se o atributo utilizado para identificar o beneficiário nos dados carregados. Este atributo é utilizado para isolar os eventos em ordem cronológica (por beneficiário) e para a exibição dos dados relativos às transações filtradas de acordo com o padrão de associação sendo analisado. Está diretamente relacionado com o cenário da pesquisa, representando o pivô sobre o qual os eventos ocorrem,

RF03 – Gerar lista classificada de acordo com a frequência dos eventos da base de dados: O modelo deve gerar uma lista ordenada de acordo com a frequência dos eventos na base de dados e de acordo com um evento selecionado pelo usuário.

- a) O usuário seleciona um evento para análise;
- b) O modelo apresenta uma lista de eventos coocorrentes com sua frequência absoluta e em ordem decrescente.

RF04 – Obter padrões de associação, cronologia e janelas de tempo: O modelo deve gerar os arquivos de entrada e executar os seguintes algoritmos de descoberta de padrões, gravando seus resultados na base de conhecimento:

- a) Apriori (BORGELT, 2010), para a geração das regras de associação. Para tanto, o modelo deve solicitar os valores de suporte e confiança mínimos a serem informados como parâmetros de execução. As regras redundantes devem ser removidas, de maneira a manter a reduzir o número de regras a serem analisadas pelo especialista;
- b) Chrono_Assoc (GOMES e CARVALHO, 2011), para a verificação do suporte e da confiança das associações sequenciais;
- c) AssocTemp (SOKOLOSKI, CARVALHO e DALLAGASSA, 2014), para a obtenção dos padrões de sequência em intervalos de tempo entre os eventos da base de dados. Para tanto, o modelo solicita um lista de eventos sobre o quais o especialista tem interesse, que será passada como parâmetro de execução do algoritmo.

RF05 - Navegar nos padrões obtidos: O modelo inicia a navegação apresentando a lista com os eventos de interesse, informada pelo especialista durante as cargas de dados, permitindo que o especialista selecione um dos eventos para exibição dos padrões obtidos, conforme as seguintes alternativas:

- a) Exibir uma lista com os eventos que associados ao evento selecionado
- b) Exibir padrões com o evento selecionado no consequente
- c) Exibir padrões com o evento no antecedente
- d) Exibir padrões com o evento tanto no consequente quanto no antecedente
- e) Exibir todas as regras, de acordo com as condições “a”, “b” ou “c”;
- f) Exibir as regras cujos suportes cronológicos sejam maiores que zero, indicando sequências na base de dados e de acordo com as condições “a”, “b” ou “c”.

Para cada regra de associação exibida, o modelo deve demonstrar a associação sequencial (GOMES e CARVALHO, 2011) e os padrões de intervalos de tempo (SOKOLOSKI, CARVALHO e DALLAGASSA, 2014) constituindo os padrões que potencialmente representam as causas e efeitos dos eventos de alto custo.

Opcionalmente, o especialista pode definir que as regras exibidas sejam aquelas que não contem os eventos selecionados, alterando o conjunto de padrões para análise.

A lista de regras pode ser filtrada, de acordo com os eventos tanto no consequente quanto no antecedente das regras. Assim, o especialista reduz o

número de regras, detalhando os eventos pelos quais tem interesse para sua análise.

RF06 - Exibir os dados relativos aos beneficiários na base de dados que estão cobertos pela regra de associação selecionada: O modelo deve permitir que o especialista selecione uma das regras de associação e, então, exibir os dados dos beneficiários que satisfaçam tal regra, de acordo com os eventos identificados na regra.

3.2.1 Arquitetura do modelo proposto

Apresenta-se o modelo proposto nesta pesquisa, com os elementos e respectiva lógica de acesso às informações e padrões, de forma a apoiar o uso rotineiro dos resultados do KDD.

Um arquivo de dados, contendo os eventos e as datas de liberação para cada beneficiário da operadora, serve de entrada para o modelo, que obtém os padrões de regras de associação com o algoritmo Apriori (AGRAWAL e SRIKANT, 1994) e implementado no programa de mesmo nome – Apriori.exe – (BORGELT, 2010), os padrões cronológicos com o Chrono_Assoc (GOMES e CARVALHO, 2011). O algoritmo AssocTemp (SOKOLOSKI, CARVALHO e DLAGASSA, 2014) é acionado, utilizando tanto o arquivo de dados quanto os eventos de interesse. Os padrões são gravados na base de conhecimento do modelo e ficam disponíveis para análise pelos especialistas (Figura 6).

O modelo conta com uma arquitetura voltada para serviços, na qual as responsabilidades do software são isoladas em diferentes componentes, enfatizando o encapsulamento, a alta coesão e o baixo acoplamento das classes implementadas.

Estão definidos dois serviços de alto nível, Dados e Padrões cujas responsabilidades são acessos aos dados, tanto para leitura quanto para gravação, carga e acesso aos padrões obtidos, respectivamente.

Além dos dois serviços de alto nível, o modelo conta com a interface de usuário e com um banco de dados, para o armazenamento dos dados de entrada, bem como os padrões e demais parâmetros para funcionamento (Figura 7).

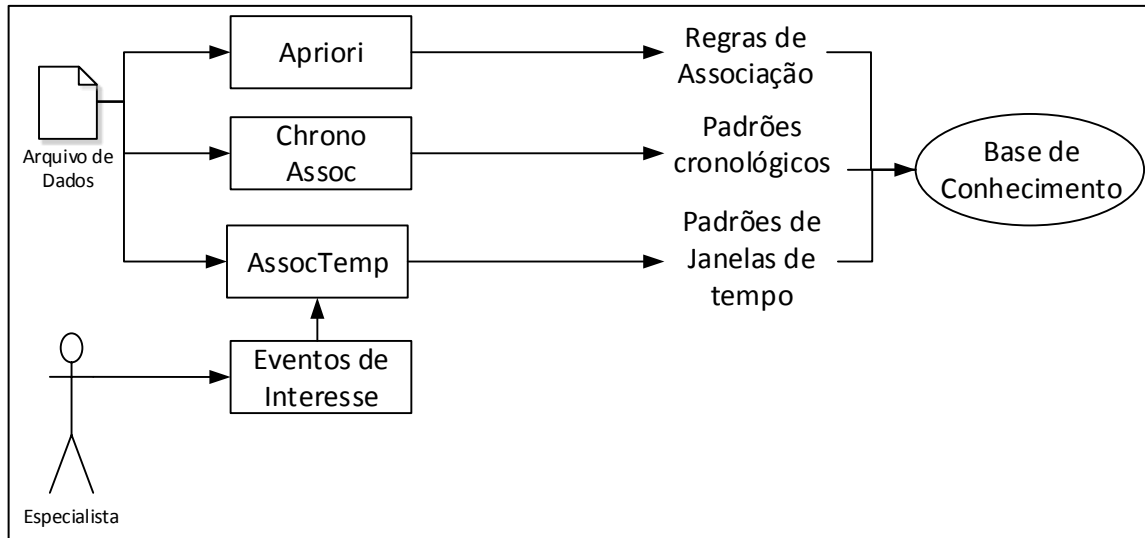


Figura 6 Representação do modelo proposto
Fonte: O autor (2015)

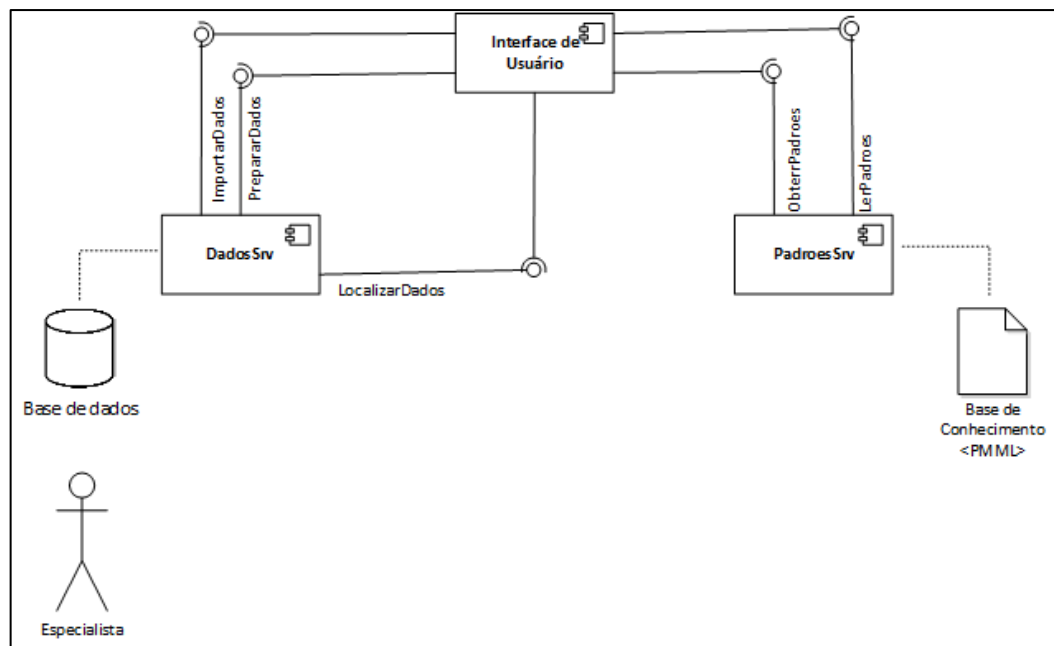


Figura 7 Arquitetura geral do modelo proposto
Fonte: O autor (2015)

Interface de usuário: Janela para navegação nos padrões obtidos. Tal interação se dará mediante a apresentação dos padrões obtidos à medida que o usuário selecione o evento pelo qual tem interesse.

Componente Dados Sua finalidade é encapsular a lógica para acesso aos dados do modelo. É responsável por carregar os dados, preparando-os para os algoritmos, bem como por localizar os dados transacionais que satisfaçam a um

determinado padrão e ainda gerar as frequências de ocorrência dos eventos na base de dados.

O componente Dados é composto pelas seguintes interfaces:

- **Dados. Importar Dados:** Importa os dados relativos às liberações de procedimentos da operadora, que contém os eventos a serem minerados acompanhados de suas datas de ocorrência, dados relativos a custos e ao local onde o procedimento foi realizado. Esta interface recebe um arquivo do tipo texto, separado por ponto-e-vírgulas e cuja primeira linha contém os nomes dos atributos e as demais linhas contém os valores dos atributos, sendo que cada linha representa uma transação diferente.
- **Dados. Preparar Dados:** Registra qual é o atributo que contém o evento a ser minerado, qual é o atributo que contém a data de ocorrência do evento e qual é o atributo que contém o elemento de negócios sobre o qual o evento ocorre. Também registra os eventos pelos quais o usuário possui interesse em investigar padrões. Ao final, gera os arquivos de dados necessários para rodar os algoritmos de Mineração de Dados.
- **Dados. Localizar Dados:** Retorna as transações que satisfazem a uma determinada regra selecionada pelo usuário na Interface de Usuário. Este elemento conecta-se ao arquivo de dados informado originalmente ao modelo e no qual estão contidas as transações com os eventos minerados.

Componente Padrões: Sua finalidade é encapsular toda a lógica que ativará cada um dos algoritmos de mineração de dados e de pós-processamento, armazenando os padrões obtidos no banco de dados, como uma base de conhecimento.

O componente Padrões é composto pelas seguintes interfaces:

- **Padrões. Obter Padrões:** Recebe os parâmetros de Suporte e Confiança mínimos e executa para executar o Apriori. Recebe os

eventos de interesse e as janelas de tempo para descoberta dos padrões temporais. Executa o Apriori e gera o arquivo PMML para as regras de associação e, na sequência, remove as regras redundantes. Executa o Chrono_Assoc e atualiza o arquivo da base de conhecimento. Executa o algoritmo de Janelas de Tempo de atualiza a base de conhecimento.

- **Padrões. Ler Padrões:** Retorna um conjunto de padrões armazenados na base de conhecimento de acordo com eventos passados como critérios de filtragem.

A integração entre os componentes do modelo deve se dar mediante chamadas por métodos implementados em suas interfaces (Figura 7), facilitando a implementação de novas funcionalidades.

De forma geral, o modelo proposto realiza duas etapas. Na primeira etapa os dados são importados e os padrões são obtidos e gravados em sua base de conhecimento. Na segunda etapa os padrões são consultados, conforme critérios do especialista e demonstrados para análise e uso.

3.3 Preparação da base de conhecimento para avaliação do modelo

Com o intuito de avaliar o modelo proposto, a base de conhecimento do modelo foi previamente carregada, permitindo que especialistas naveguem nos padrões. Para tanto, foram importados 16.355 registros de liberações de procedimentos relativos a 127 beneficiários e contendo 2.083 procedimentos realizados entre agosto de 2008 e outubro de 2014.

A base de conhecimento foi carregada com 12.992 regras de associação com suporte mínimo 5%, de maneira que associações menos frequentes sejam obtidas, e confiança mínima de 80%, buscando regras cuja associação sejam fortes e, portanto, relevantes. As regras de associação foram pós-processadas com o Chrono_Assoc, resultando em 12.305 regras de associação sequenciais, restando 687 regras de associação não sequenciais.

O especialista indicou 8 procedimentos de interesse para obtenção das janelas de tempo, levando em conta intervalos de 90, 120 e 150 dias, resultando em 5.554 padrões. (Quadro 7).

ANTÍGENO CARCINOGENICO (CEA), DOSAGEM
GASOMETRIA (PH, PCO2, SA, O2, EXCESSO BASE)
TRAX 1 – INCIDENCIA
MARCADORES TUMORAIS (CA 19.9, CA 125, CA 72-4, CA 15-3), DOSAGEM
CONSULTA EM PRONTO SOCORRO
ÁCIDO LÁTICO, DOSAGEM

Quadro 7 Procedimentos de interesse indicado pelo especialista
 Fonte: O autor (2015)

3.4 Avaliação do modelo proposto

A fim de avaliar o modelo proposto e o seu potencial para uso na rotina em Saúde, será aplicado um instrumento de avaliação que permita aos especialistas demonstrarem o quanto o modelo é adequado para uso em suas rotinas.

A avaliação, bem como o instrumento de avaliação, aborda quatro eixos principais, ou construtos que refletem as dimensões do modelo a serem avaliadas pelos especialistas envolvidos (Quadro 8).

Construto	Questões
Satisfação do especialista quanto ao modelo	Q1. O sistema mostra informações claramente e é fácil compreendê-las
	Q2. Aprendi rapidamente a utilizar a interface de navegação do sistema
	Q3. Utilizaria o sistema em minha rotina de trabalho
Apresentação dos padrões e informações	Q4. A frequência dos eventos é útil para a análise dos padrões
	Q5. Os padrões apresentados são claros e facilmente compreendidos
	Q6. Os dados apresentados sobre os beneficiários são úteis e relevantes para meu trabalho
	Q7. A disposição dos padrões e informações é adequada para o uso em minha rotina
Capacidade de descrever os eventos	Q8. Os padrões apresentados esclarecem como os eventos estão relacionados.
	Q9. As janelas de tempo permitem compreender como os eventos estão relacionados.
	Q10. Compreendi claramente os padrões apresentados e como eles descrevem os eventos ocorridos.
Perfil dos entrevistados	Q11. Sua área de formação/especialidade
	Q12. Há quanto tempo atua com auditoria?
	Q13. Utiliza ferramentas estatísticas para a análise de procedimentos a serem auditados?

Quadro 8 Construtos e respectivas questões para a avaliação do modelo proposto
Fonte: O autor (2015)

As repostas possíveis consideraram a escala psicométrica de Likert (1932), com três possibilidades variando desde a negação total até a aceitação total da informação, no qual cada resposta recebe um peso relativo à sua posição na escala de 0 a 2:

- a) **Discordo**, peso = 0;
- b) **Sou Indiferente**, peso = 1;
- c) **Concordo**, peso = 2.

O construto Perfil dos entrevistados, contará com respostas em escala diversa à escala Likert, conforme exposto no Quadro 9.

Questão	Chave de respostas
Q11. Área de formação/especialidade	Resposta aberta.
Q12. Há quanto anos atua com auditoria?	_____ anos.
Q13. Utiliza ferramentas estatísticas para a análise de procedimentos a serem auditados?	() Sim () Não Se respondeu não, descreva como realiza a análise de procedimentos.

Quadro 9 Chaves de respostas para as questões relativas ao perfil dos entrevistados
Fonte: O autor (2015)

A aplicação da avaliação se dará mediante a experimentação do modelo construído e a aplicação de formulário de pesquisa (Apêndice C) em sessão coletiva, envolvendo 6 especialistas em auditoria de procedimentos de operadora de Saúde e terá duração de 1,5 horas distribuídas em 15 minutos para apresentação dos conceitos inseridos no modelo proposto, 15 minutos para explanação dos objetivos, 45 minutos para experimentação e 15 minutos para avaliação do modelo.

Para tanto, será utilizado um conjunto de dados selecionados pelos especialistas para a obtenção dos padrões diferente daquela utilizada para o levantamento de requisitos (Etapa 2 do método), contendo 16.355 liberações de procedimentos realizados entre agosto de 2008 e outubro de 2014 relativas a 127 e beneficiários e contendo 2.083 procedimentos. Os dados serão carregados e os padrões e demais informações serão gerados pelo modelo, em momento anterior à avaliação, de forma a testar a capacidade de geração e armazenamento.

Os especialistas receberão um manual de uso, explicando o funcionamento geral do modelo, de forma que se familiarizem com a interface e com o uso.

A avaliação da aceitação do modelo proposto será realizada mediante a tabulação das respostas e da determinação do índice de validade de conteúdo (IVC), proposto por Waltz *et al* (1991, *apud* DALLAGASSA, 2009) conforme o seguinte processo:

1. Somar os pesos obtidos em cada construto, com base nas respostas dos especialistas;
2. O total máximo possível para cada construto é obtido somando os máximos pesos de cada resposta relativa ao construto em questão;

3. O IVC é obtido dividindo-se o total obtido pelo total máximo possível, multiplicado por 100 (Expressão 10).

$$IVC_{Construto} = \left(\frac{Total\ obtido_{Construto}}{Total\ máximo\ possível_{Construto}} \right) 100 \quad (10)$$

O modelo proposto também foi avaliado mediante o IVC, da mesma forma que os construtos. Somando-se os totais obtidos de todos os construtos, dividindo pela soma dos totais máximos possíveis e multiplicando o quociente por 100.

Para fins de avaliação, valores obtidos abaixo de 70% são considerados como não aceitos, entre 70% e 80% são considerados como parcialmente aceitos e acima de 80% são considerados totalmente aceitos.

3.5 Aspectos éticos

Por questões éticas, foram omitidos todos os dados relativos à identificação dos beneficiários cujos procedimentos foram utilizados para a obtenção dos padrões na base dados. Os números de identificação dos beneficiários foram substituídos por valores alterados diretamente no arquivo de transações, sem que o autor ou quaisquer indivíduos envolvidos na pesquisa tenham acesso a tais informações.

Ainda, os especialistas envolvidos na avaliação do modelo firmaram Termo de Consentimento Livre e Esclarecido, declarando ciência da pesquisa e concordando com os termos da avaliação do modelo (Apêndice E).

O projeto foi submetido e autorizado pelo comitê de Ética em Pesquisa da PUC-PR, porque os dados foram obtidos em parceria com tal instituição, sob nº 864.236, anexo I.

4 APRESENTAÇÃO E DISCUSSÃO DOS RESULTADOS

Esta seção apresenta os resultados observando as etapas propostas para a pesquisa. São apresentados os pontos de atenção e os requisitos dos especialistas adotados no modelo proposto, com o intuito de aumentar as chances de seu uso na rotina em gestão da Saúde.

O modelo proposto é detalhado, de forma a demonstrar como está organizado e como seu funcionamento é viabilizado.

A leitura e interpretação dos padrões obtidos pelo modelo proposto é apresentada mediante quatro padrões descobertos, exemplificando como os padrões, podem ser explorados pelos especialistas.

No processo de avaliação do modelo proposto estão apresentados os resultados observando os critérios propostos para tal finalidade.

4.1 Pontos de atenção para adoção do kdd de dados na saúde

Analisando as 18 publicações selecionadas por Carvalho et al (2014), percebe-se que 15 estudos (83%) relatam experimentos envolvendo a área clínica e 5 estudos (27%) envolvendo gestão da saúde, dos quais 3 envolvem somente gestão e 2 abordam tanto a questão clínica quanto a questão gerencial. Destaca-se que Sobre as três grandes etapas do KDD, todos os autores relatam Mineração de Dados, 78% também discutem a etapa de Pré-Processamento e apenas 39% o Pós-processamento. Além de mostrar que o Pós-processamento é menos prevalente, sugere que o cuidado em depurar e demonstrar os padrões de maior potencial para os especialistas ainda é uma área menos explorada e que merece maior atenção.

No caso desta pesquisa, o pós-processamento dos padrões é fundamental, pois o Ponto de Atenção Fornecimento de subsídios para explicações causais somente foi possível utilizando-se técnicas para associação sequencial, para o que foram utilizados o Chrono_Assoc (GOMES E CARVALHO, 2011) e o AssocTemp (SOKOLOSKI, CARVALHO e DALLAGASSA, 2014).

No sentido de apresentar e compreender como os pontos de atenção podem ser combinados, conforme a pesquisa, segue relato das associações obtidas, que estão integralmente representadas no Apêndice D:

Muito embora a Representação Visual dos Resultados não seja citada por 12 dos 18 estudos analisados, as regras de exceção obtidas demonstram que este ponto de atenção está presente quando Fornecimento de subsídios para explicações

causais, Detecção de padrões em tempo real e Auxílio ao planejamento em saúde são citados pelos autores. (Quadro 10)

Ponto de Atenção	Pontos de Atenção Associados
Representação visual dos resultados	
	Auxílio ao planejamento em saúde
	Detecção de padrões em tempo real
	Explicações causais

Quadro 10 Pontos de atenção associados com Representação visual dos resultados.
Fonte: O autor (2015)

A associação entre Representação Visual dos Resultados e Auxílio ao planejamento em saúde, bem como a Explicações causais levam a hipótese de que a usabilidade das ferramentas que explorem o KDD para apoio à decisão é um elemento a ser considerado e pode colaborar com a adoção da Mineração dos Dados na rotina em Saúde, pois agrega à capacidade de descrever eventos e de planejar ações por parte do especialista.

O ponto de atenção Descrição de eventos ocorridos foi citado por 6 autores estudados e está presente em 10 regras gerais, estando ausente em 5 regras gerais.

Entretanto os pontos de atenção Desenvolvimento de funcionalidade específica, Avaliação (subjetiva ou objetiva) da relevância dos padrões encontrados, Representação visual dos resultados e Obediência a processo de trabalho específico levam à presença do ponto Descrição de eventos ocorridos, obtidos mediante o uso de regras de exceção (Quadro 11).

Ponto de Atenção	Pontos de Atenção Associados
Descrição de eventos	Avaliação da relevância dos padrões encontrados
	Funcionalidade específica
	Processo específico
	Representação visual dos resultados

Quadro 11 Pontos de atenção associados com Descrição de eventos ocorridos.
Fonte: O autor (2015)

Já o ponto de atenção Desenvolvimento de funcionalidade específica é citado por 9 autores, colocando-o entre os três pontos de atenção mais citados.

Entretanto, foram obtidas 14 regras gerais, das quais uma conta com a presença do ponto Desenvolvimento de funcionalidade específica.

Dentre as regras gerais em que o Desenvolvimento de funcionalidade específica está ausente, as regras de exceção mostram, por sua vez, que Detecção de padrões em tempo real, Obediência a processo de trabalho específico, Combinação de diferentes tipos de tarefas de Mineração de Dados, Utilização de modelo de informações específicas e Descrição de eventos ocorridos numa determinada população levam à presença do ponto de atenção Desenvolvimento de funcionalidade específica (Quadro 12).

Ponto de Atenção	Pontos de Atenção Associados
Funcionalidade específica	Descrição de eventos ocorridos
	Detecção de padrões em tempo real
	Diferentes tarefas de Mineração de Dados
	Modelo de informações específicas
	Processo específico

Quadro 12 Pontos de atenção associados com Desenvolvimento de funcionalidade específica.

Fonte: O autor (2015)

Esta constatação leva a considerar que o Desenvolvimento de funcionalidade específica é menos prevalente quando se trata de apoio o planejamento em saúde e que esta se torna uma condição favorável para a pesquisa, dado que o modelo aqui proposto pode apoiar o desenvolvimento de funcionalidades específicas para a exploração de padrões obtidos durante o KDD e que explora padrões relativos à ocorrência de eventos, abrindo espaço para pesquisas futuras.

O ponto de atenção Fornecimento de subsídios para explicações causais é citado por 8 autores, ficando entre os 4 pontos de atenção mais citados. Foram obtidas 17 regras gerais, dentre as quais, 10 não contam com o ponto de atenção Fornecimento de subsídios para explicações causais.

As regras de exceção, no entanto, mostram que Auxílio ao planejamento em saúde, Avaliação (subjetiva ou objetiva) da relevância dos padrões encontrados, Detecção de padrões em tempo real, Combinação de diferentes tipos de tarefas de Mineração de Dados, Utilização de modelo de informações específicas, Previsão da ocorrência de eventos numa determinada população, Obediência a processo de

trabalho específico, Apropriação de protocolo específico, Representação visual dos resultados levam à presença do ponto de atenção Fornecimento de subsídios para explicações causais (Quadro 13).

Ponto de Atenção	Pontos de Atenção Associados
Explicações causais	Apropriação de protocolo específico
	Auxílio ao planejamento em saúde
	Avaliação da relevância dos padrões encontrados
	Detecção de padrões em tempo real
	Diferentes tarefas de Mineração de Dados
	Modelo de informações específicas
	Previsão de eventos
	Processo específicos
	Representação visual dos resultados

Quadro 13 Pontos de atenção associados com Fornecimento de subsídios para explicações causais.

Fonte: O autor (2015)

4.2 Requisitos dos especialistas

Os requisitos levantados junto aos especialistas da Operadora de Saúde (Quadro 14) evidenciam a necessidade de se navegar nas regras de associação obtidas, demonstrando quais são os padrões pós-processados (associação cronológica e janelas de tempo) relacionados com tal regra. As diferentes filtragens indicadas como requisitos pelos especialistas mostram a necessidade de se reduzir o volume de regras a serem manipuladas durante as análises, o que pode contribuir para a eficiência do processo.

A solicitação de se exibir os dados dos beneficiários cobertos pelos padrões selecionados mostra a necessidade de relacionamento entre os padrões obtidos e os dados transacionais, permitindo aos analistas a identificação das transações envolvidas no padrão e suportando a ação direta sobre tais situações.

Req Nº	Resumo	Justificativa do especialista
1	Adicionar janelas de tempo de 90,120 e 150 dias.	A auditoria tem como padrão analisar dados com intervalos de 150 dias. Entretanto, a análise dos padrões com intervalos de 120 dias proposta no experimento inicial levou a necessidade de considerar 90 e 120 dias também, de maneira que o intervalo de análise seja estendido e permita-se identificar possíveis janelas que não estão consideradas nos 150 dias. Estes intervalos foram considerados em reunião com os especialistas. Assim a comparação com as demais janelas de tempo propostas pode indicar a necessidade de ajuste nos intervalos analisados.
2	Permitir a filtragem dos procedimentos tanto no antecedente quanto no consequente das regras.	Isto permite a análise da ocorrência de procedimentos de acordo com a necessidade do analista, reduzindo o número de padrões a serem analisados.
3	Exibir os dados dos beneficiários e respectivas fichas médicas (importados para o modelo) que atendem a regra selecionada para análise.	Permitir a verificação de quais beneficiários e liberação de procedimentos, bem como seus custos, estão cobertos pela regra selecionada.
4	Filtrar as regras que não possuem o procedimento selecionado	Permitir a análise das regras que não estão associadas como o procedimento em questão.
5	Permitir a filtragem dos dados dos beneficiários e respectivas fichas médica que não são cobertos pela regra selecionada.	Analisar os dados relativos aos beneficiários e procedimentos bem como seus custos, que não satisfazem a regra selecionada.
6	Gerar frequência dos procedimentos concorrentes com procedimento selecionado pelo especialista.	Exibir frequência dos procedimentos registrados na base de dados indicando quais são mais comuns.

Quadro 14 Requisitos e justificativas apontadas pelos especialistas em apresentação preliminar

Fonte: O autor (2015)

A exploração dos requisitos dos especialistas permite a organização de funcionalidades práticas, voltadas para as necessidades reais e que podem ser fundamentais para a adoção cotidiana dos padrões obtidos com o KDD. -

4.3 Modelo proposto

Apresenta-se o modelo proposto desenvolvido nesta pesquisa, com suas estruturas e respectivo protótipo. Tal modelo foi desenvolvido levando em conta tanto os pontos de atenção para o uso da Mineração de Dados na Saúde quanto os requisitos levantados junto aos especialistas.

A lógica de acesso a dados, a obtenção dos padrões e a interface de navegação são descritos de forma demonstrar a sua construção.

4.3.1 Importação de dados e geração da base de conhecimento

O modelo requer a entrada inicial de um arquivo do tipo texto separado por vírgulas contendo os dados relativos aos procedimentos liberados para os beneficiários, conforme o formato expresso no Quadro 15.

Posição	Atributo	Descrição
1	codigo_beneficiário	Número identificador único do beneficiário originado na base de dados de liberações da operadora.
2	Sexo	Gênero Masculino (M) ou Feminino (F) do beneficiário.
3	Idade	Número representando a idade do beneficiário.
4	cidade	Município de residência o de realização do procedimento do beneficiário.
5	data_atendimento	Dia, mês e ano de realização do procedimento.
6	codigo_servico	Identificador único do serviço ou procedimento realizado originado na base de dados da operadora.
7	tipo_servico	Identificado único da categoria do serviço ou procedimento realizado originado na base de dados da operadora.
8	servico	Texto descritivo do serviço ou procedimento realizado originado na base de dados da operadora.
9	codigo_cid	Código Internacional da Doença definido pelo médico ou pela operadora.
10	nome_prestador	Nome do médico ou parceiro solicitante do serviço ou procedimento realizado originado na base de dados da operadora.
11	local_atendimento	Município no qual o serviço ou procedimento foi realizado. Originado na base de dados da operadora.

12	quantidade	Número indicando a quantidade de serviços ou procedimentos foram liberados ou realizados.
13	Custo	Valor numérico indicando o custo do serviço ou procedimento realizado. Originado na base de dados da operadora.

Quadro 15 Layout do arquivo de entrada com os procedimentos liberados pela operadora de saúde

Fonte: O autor (2015)

Esse arquivo de entrada é a implementação do arquivo dado na representação do modelo proposto na Figura 7, seção 3.2.1 Arquitetura do modelo proposto e é gerado a partir das bases de dados contendo as transações da operadora de Saúde, sendo de responsabilidade do especialista.

O modelo lê o arquivo de procedimentos e o particiona em um conjunto de dados sobre as pessoas (beneficiários) e outro conjunto de dados sobre os procedimentos, gravando ambos os conjuntos em sua base de dados.

O modelo também requer a importação de um arquivo texto contendo apenas os códigos de identificação dos eventos de interesse para análise, conforme layout demonstrado na Figura 8. Para tanto, o modelo marca os procedimentos em sua base de dados como procedimentos de interesse e, a partir disso, gera uma lista de procedimentos reduzida (interesse), que será exibida durante a navegação do especialista.

Os procedimentos marcados como “de interesse” são utilizados para a obtenção dos padrões de janelas de tempo, servindo como parâmetro de execução do AssocTemp.

10101039AMB
40301109AMB
40316122AMB
40302016AMB
92403034MED
40316378AMB
40304507AMB
40805018AMB

Figura 8 Fragmento de arquivo com procedimentos de interesse, a ser importado pelo modelo

Fonte: O autor (2015)

A partir da importação inicial dos arquivos de dados, o modelo automatiza a obtenção das regras de associação com Apriori (BORGELT, 2010), pós processa os padrões obtendo regras sequenciais com o Chrono_Assoc (GOMES E CARVALHO, 2011) e, também obtém as janelas de tempo com o AssocTemp (SOKOLOSKI, CARVALHO E DALLAGASSA, 2014), mediante a geração de arquivos específicos e a execução de cada algoritmo. A sequência de processamento do modelo, bem como os arquivos de entrada e saída são demonstrados na Figura 9.

Para a obtenção dos padrões, o modelo gera dois arquivos de saída a partir dos dados importados para sua base. O arquivo Eventos contem os procedimentos liberados para os beneficiários e é utilizado como entrada para o Apriori. O arquivo Eventos-datas contém os mesmos procedimentos, mas acompanhados das datas de realização, de forma a servir de entrada para o AssocTemp (Figura 10).

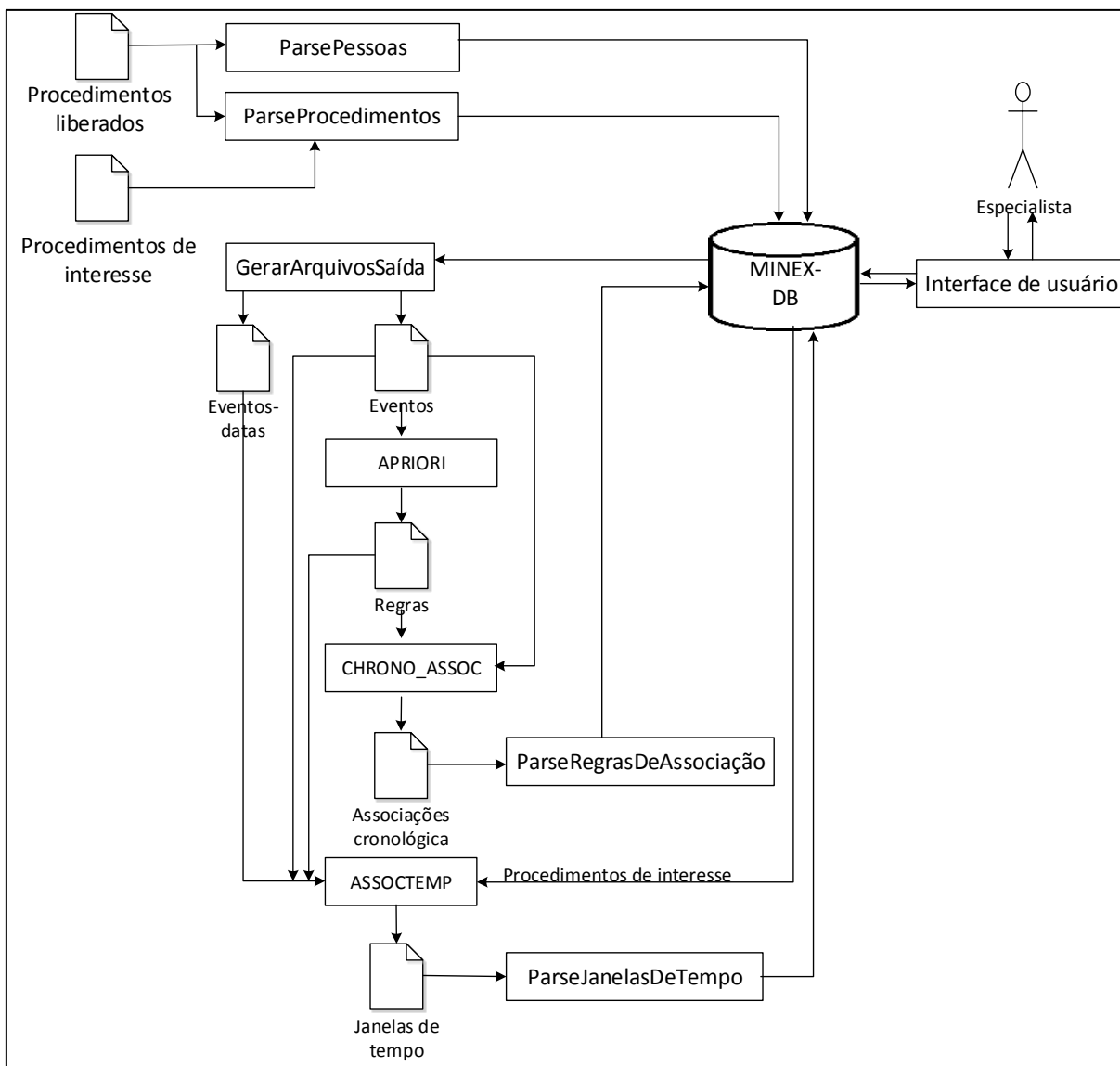


Figura 9 Sequência de arquivos gerados e processamento para obtenção dos padrões sobre procedimentos de alto custo
Fonte: O autor (2015)

Arquivo para apriori						
41001079AMB	41001079AMB	41001079AMB	41001079AMB	41001079AMB	74683624MAT	
20103522AMB	20103522AMB	20103522AMB	20103522AMB	20103522AMB	10101012AMB	
94249911MED	40901106AMB	10101012AMB				
74816543MAT	80017010HOS					
Arquivo para AssocTemp						
41001079AMB(26/06/2014),41001079AMB(26/06/2014),41001079AMB(26/06/2014),41001079AMB(26/06/2014),41001079AMB(26/06/2014),74683624MAT(26/06/2014),20103522AMB(03/07/2014),20103522AMB(04/07/2014),20103522AMB(07/07/2014),20103522AMB(08/07/2014),20103522AMB(11/07/2014),10101012AMB(18/08/2014),94249911MED(03/07/2014),40901106AMB(08/07/2014),10101012AMB(07/08/2014)						
74816543MAT(06/01/2014),80017010HOS(10/03/2014)						

Figura 10 Fragmentos dos arquivos de entrada para Apriori e AssocTemp gerados pelo modelo
Fonte: O autor (2015)

O modelo executa o Apriori, passando como parâmetro o arquivo Eventos, resultando no arquivo Regras, contendo as regras de associação obtidas (Figura 11).

40301583AMB	<-	40301605AMB	40302547AMB	(10.2%, 92.3%)
40301583AMB	<-	40301605AMB	40302040AMB	(10.2%, 92.3%)
40301583AMB	<-	40301605AMB	40302504AMB	(10.2%, 92.3%)
40301583AMB	<-	40301605AMB	40301630AMB	(10.2%, 92.3%)
40301605AMB	<-	40301583AMB	40301630AMB	(10.2%, 92.3%)
40304361AMB	<-	40301605AMB	40301583AMB	(11.0%, 92.9%)
40301583AMB	<-	40301605AMB	40304361AMB	(11.0%, 92.9%)
40301605AMB	<-	40301583AMB	40304361AMB	(10.2%, 100.0%)

Figura 11 Fragmento de arquivo gerado pelo Apriori durante a criação da base de conhecimento pelo modelo
Fonte: O autor (2015)

As associações cronológicas são obtidas pelo modelo mediante a execução do programa Chrono_Assoc, passando como entradas os arquivos Eventos e Regras. Ao final, o Chrono_Assoc grava um arquivo com o resultado do pós-processamento (Figura 12).

40301583AMB	<-	40301605AMB	40302547AMB	(10.2%, 92.3%)	(6.29921%, 66.6667%)
40301583AMB	<-	40301605AMB	40302040AMB	(10.2%, 92.3%)	(3.14961%, 33.3333%)
40301583AMB	<-	40301605AMB	40302504AMB	(10.2%, 92.3%)	(3.93701%, 41.6667%)
40301583AMB	<-	40301605AMB	40301630AMB	(10.2%, 92.3%)	(3.93701%, 41.6667%)
40301605AMB	<-	40301583AMB	40301630AMB	(10.2%, 92.3%)	(0.787402%, 8.33333%)
40304361AMB	<-	40301605AMB	40301583AMB	(11.0%, 92.9%)	(7.87402%, 76.9231%)
40301583AMB	<-	40301605AMB	40304361AMB	(11.0%, 92.9%)	(2.3622%, 23.0769%)
40301605AMB	<-	40301583AMB	40304361AMB	(10.2%, 100.0%)	(0%, 0%)

Figura 12 Fragmento de arquivo gerado pelo Chrono_Assoc durante a criação da base de conhecimento pelo modelo
Fonte: O autor (2015)

Os padrões de intervalo de tempo são obtidos mediante o pós-processamento do arquivo Regras com o programa AssocTemp (SOKOLOSKI, CARVALHO E DALLAGASSA, 2014), passando como entradas o arquivo Eventos-Datas e os procedimentos de interesse definidos pelo especialista e gravados na base de dados.

Ao final, é gravado um arquivo para cada evento de interesse em função de cada intervalo de tempo a ser analisado. Assim, considerando os eventos foram analisados três intervalos de tempo, 90, 120 e 150 dias para permitir uma análise mais ampla do comportamento temporal das associações, e que a análise deve ocorrer quando o evento está tanto no antecedente quanto no consequente da regra,

seis arquivos foram gerados pelo AssocTemp para cada evento de interesse (Figura 13).

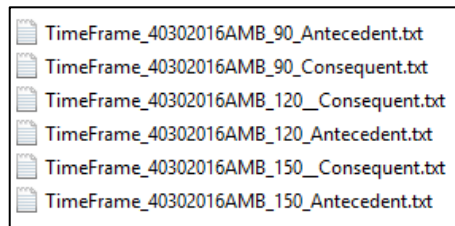


Figura 13 Exemplo de arquivos com padrões de intervalo de tempo gerados para um procedimento de interesse pelo modelo

Fonte: O autor (2015)

Cada saída gerada pelo AssocTemp, cujo formato é exemplificado na Figura 14, é lida e gravada na base de conhecimento do modelo.

```
Arquivo de entrada (base de dados): events.txt
Arquivo de entrada com regras: rules.txt
Item selecionado: 40805018AMB
Limite máximo: 150
Número total de regras na entrada: 12992
Número total de regras na saída: 140

-----
Layout de saída

Consequente <- Antecedente(PERC1, PERC2) (PERC3, PERC4) (PERC5, PERC6) (PERC7)

PERC1 = # Antecedente / # Total de registros
PERC2 = # Registros com Antecedente e Consequente / # Antecedente
PERC3 = # Primeira ocorrência antecedente e consequente / # Total de registros
PERC4 = # Primeira ocorrência antecedente e consequente dentro da janela temporal / #Primeira ocorrência antecedente e consequente
PERC5 = # Antecedente e consequente / # Total de registros
PERC6 = # Antecedente e consequente dentro da janela temporal / # Antecedente e consequente
PERC7 = # Total de vezes que a regra aconteceu pelo menos uma vez.

-----

40805018AMB <- 40805018AMB (11.719%, 100.000%) (4626.563%, 99.054%) (15.000)
40805018AMB <- 92387373MED 40302237AMB (5.5%, 85.7%) (2.3%, 100.0%) (73.4%, 100.0%) (3.0)
40805018AMB <- 92417396MED 40301400AMB (5.5%, 85.7%) (2.3%, 100.0%) (316.4%, 85.2%) (3.0)
40805018AMB <- 92417396MED 40302040AMB (5.5%, 85.7%) (2.3%, 100.0%) (86.7%, 89.2%) (3.0)
40805018AMB <- 75600110MAT 40302016AMB (5.5%, 85.7%) (2.3%, 100.0%) (7726.6%, 88.7%) (3.0)
40805018AMB <- 75600110MAT 40301397AMB (5.5%, 85.7%) (2.3%, 100.0%) (1775.0%, 89.0%) (3.0)
40805018AMB <- 75600110MAT 40302040AMB (5.5%, 85.7%) (2.3%, 100.0%) (159.4%, 92.6%) (3.0)
40805018AMB <- 75600110MAT 40302423AMB (5.5%, 85.7%) (2.3%, 100.0%) (18699.2%, 86.7%) (3.0)
40805018AMB <- 75600110MAT 40302580AMB (5.5%, 85.7%) (2.3%, 100.0%) (17960.9%, 87.2%) (3.0)
```

Figura 14 Fragmento de arquivo gerado pelo AssocTemp

Fonte: O autor (2015)

Como os especialistas indicaram, inicialmente, 8 eventos de interesse, o modelo gerou 48 arquivos diferentes, que foram lidos, convertidos e gravados na base de conhecimento. Entretanto, este número varia de acordo com os eventos de interesse e com os intervalos de tempos a serem analisados.

Os arquivos de saída do Chrono_Assoc e do AssocTemp são, então, importados para a base de dados do modelo, de acordo com o modelo de dados exposto na Figura 15.

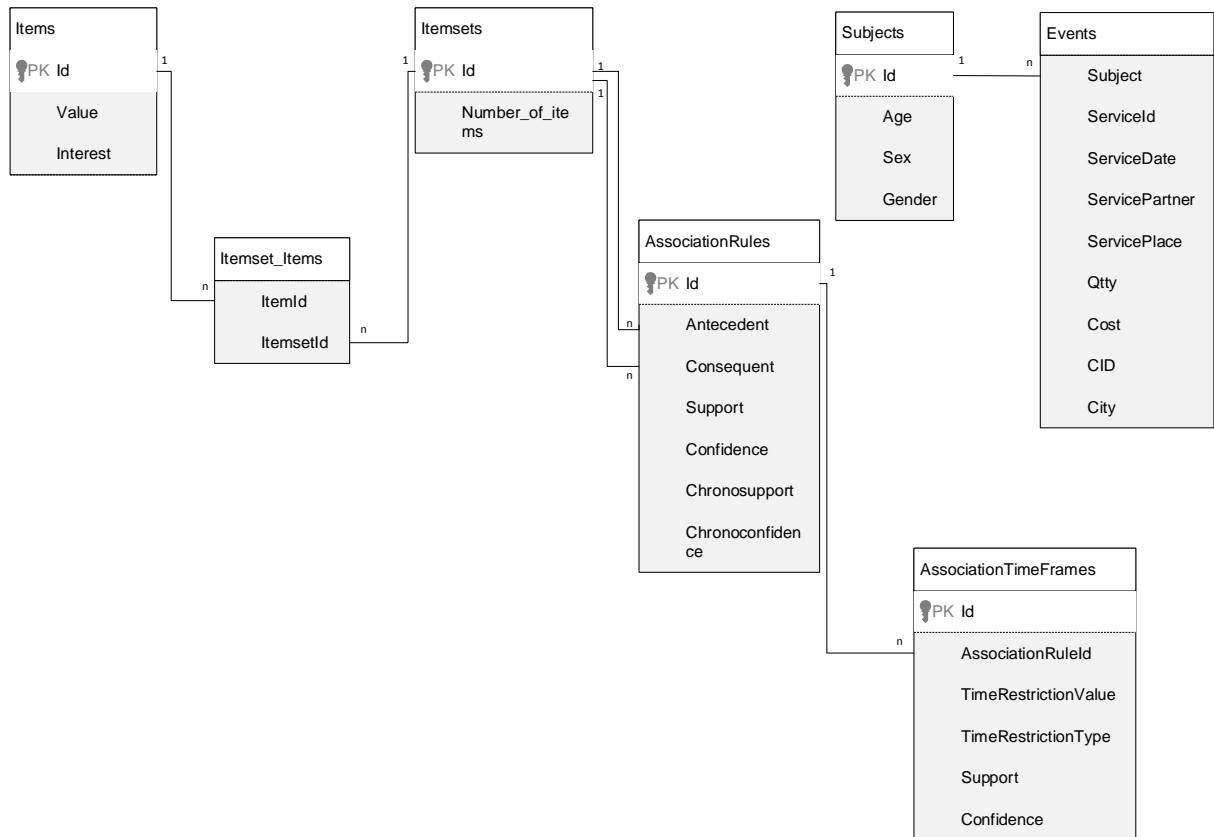


Figura 15 Modelo de entidades e relacionamentos do banco de dados do modelo proposto
Fonte: O autor (2014)

A base de dados foi elaborada de maneira a respeitar a normalização de banco de dados, permitindo consistência nos dados e melhor desempenho para o modelo proposto.

A tabela **Items** armazena os procedimentos exclusivos, obtidos a partir do arquivo de entrada fornecido pelo especialista, incluindo a indicação de procedimentos de interesse, no atributo Interest.

A tabela **ItemSets** contém um identificador exclusivo para cada conjunto de itens nas regras obtidas pelo Apriori (BORGELT, 2010) e a tabela Itemset_Items relaciona os Item aos respectivos Itemsets, garantindo que o modelo seja capaz de representar a ocorrência de múltiplos itens no antecedente das regras de associação sem, no entanto, replicar dados desnecessariamente.

A tabela **AssociationRule** armazena os dados relativos às regras de associação, onde Antecedent e Consequent são os atributos que apontam para os conjuntos de Itemsets que compõem a regra de associação. Esta tabela também armazena as métricas de suporte, confiança, suporte cronológico e confiança cronológica.

A tabela **AssociationTimeFrames** armazena os padrões de janelas de tempo para cada uma das regras de associação obtidas pelo modelo.

Os dados relativos aos beneficiários são armazenados na tabela **Subjects** e os dados relativos às liberações de procedimentos são armazenados na tabela **Events**.

A lógica de funcionamento do modelo é executada mediante uma série de classes, construídas em Java e que encapsulam os procedimentos necessários para todo o processamento requerido pelo modelo proposto (Figura 16).

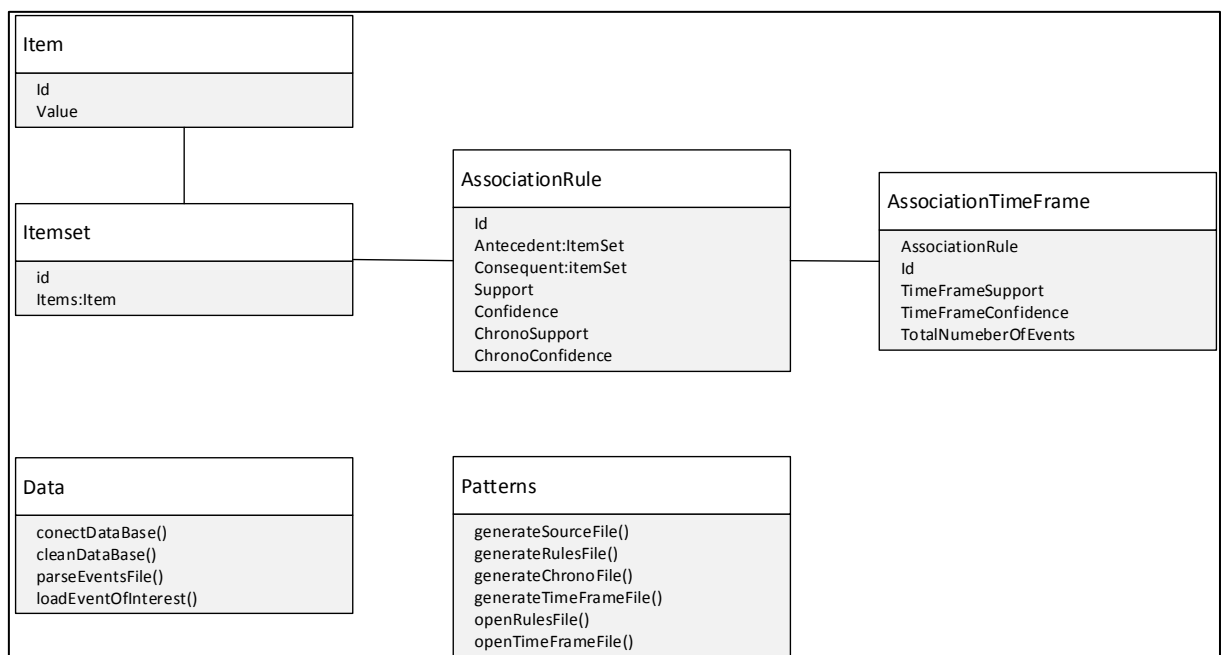


Figura 16 Classes implementadas no modelo
Fonte: O autor (2014)

Por uma questão de abstração de linguagens de programação, os atributos e métodos demonstrados na figura 17 são aqueles que estão diretamente envolvidos na lógica de funcionamento.

A classe **AssociationRules** é o principal elemento da implementação da base de conhecimento do modelo, uma vez que ela instancia as regras de associação obtidas com o Apriori, juntamente com o suporte e a confiança cronológicos obtidos com o Chrono_Assoc. Para tanto, essa classe instancia um objeto da classe Itemset para o antecedente da regra (atributo Antecent). Esse objeto Itemset, por sua vez, instancia os procedimentos (um ou muitos) determinados pelo Apriori no antecedente, como objetos da classe Item. O mesmo é realizado para instanciar o consequente da regra. Assim, garante-se que todos os

procedimentos que compõem o antecedente e o consequente da regra serão instanciados dentro de cada regra de associação no modelo proposto.

A classe **AssociationTimeFrame** instancia os padrões obtidos com os AssocTemp. Optou-se por uma associação simples, ao contrário de um atributo do tipo lista na classe AssociationRules, para permitir a navegação inversa, na qual é possível instanciar uma regra de associação (da classe AssociationRule) a partir de uma instância de um intervalo temporal (AssociationTimeFrames)

A classe Item instancia os procedimentos realizados e a classe Itemset instancia os conjuntos de itens (itemsets) presentes no antecedente (AssociationRule.Antecedent) ou no consequente (AssociationRule.Consequent) das regras de associação.

A classe **Data** implementa os métodos a serem utilizados para conexão com a base de conhecimento (connectDataBase), para limpar a base de conhecimento em caso de atualização dos padrões (cleanDataBase) e para importação do arquivo do tipo texto (parseEvents)

A classe Data também implementa o método loadEventsOfInterest, que importa o arquivo tipo texto contendo os procedimentos sobre os quais o especialista tem interesse e que serão utilizados como parâmetros para execução do AssocTemp.

A classe **Patterns** implementa os métodos para obtenção dos padrões a serem gravados na base de conhecimento.

O método **generateSourceFile** gera os arquivos que serão utilizados como entrada do Apriori, do Chrono_Assoc e do AssocTemp.

O método **generateRulesFile** executa o Apriori, passando os parâmetros necessários para a obtenção das regras de associação e informados pelo usuário no momento da criação da base de conhecimento.

O método **generateChronoFile** executa o Chrono_Assoc, passando os parâmetros necessários para o pós-processamento das regras de associação e obtenção do suporte e confiança cronológicos.

O método **generateTimeFrameFile** executa o AssocTemp, passando os parâmetros necessários para o pós-processamento das regras de associação e obtenção dos padrões de intervalo de tempo.

O método **openRuleFile** lê a saída do Chrono_Assoc e grava as regras de associação na base de dados, utilizando um método privado para conversão do arquivo nas classes Item, AssociationRule, ItemSet.

O método **OpenTimeFrameFile** lê cada um dos arquivos gerados pelo AssocTemp e grava a base de conhecimento, utilizando um método privado para conversão do arquivo na classe AssociationTimeFrame.

4.3.2 Interação do usuário com o sistema

A interação do usuário com o modelo proposto se dá em três momentos. Primeiramente, o arquivo com os procedimentos deve ser importado para a base de dados do modelo (Figura 17).

Então, os padrões devem ser descobertos, momento em que a base de conhecimento é gravada. Após a gravação da base de conhecimento, os especialistas podem navegar nos padrões a fim de explorá-lo.

A interface de usuário do modelo tem ênfase na facilidade de navegação, de maneira que seu uso seja simples e contribua para a eficiência das análises dos padrões. Assim, o modelo possui três janelas. Uma para a obtenção das regras de associação e para pós-processamento dos padrões (Figura 18); Uma janela para análise dos padrões e uma janela que exibe detalhes a ficha médica com os os procedimentos realizados por um dado beneficiário.

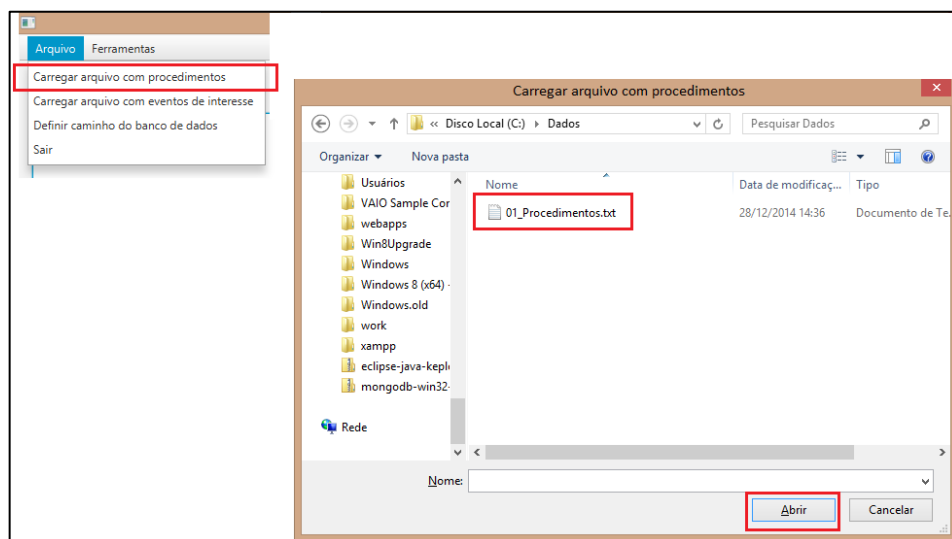


Figura 17 Caixa de diálogo Carregar arquivo com procedimentos
Fonte: O autor (2014)

Os padrões são obtidos mediante a tela “Obter Regras de Associação” (Figura 19Figura 18), na qual o usuário indica o suporte e confiança mínimos; o número mínimo e máximo de itens nas regras para o Apriori; o caminho do arquivo contendo os eventos de interesse e os intervalos de tempo para o AssocTemp.

Figura 18 Interface para geração da base de conhecimento do modelo
Fonte: O autor (2014)

Ao clicar no botão “Obter regras”, o modelo gera os arquivos de entrada para os algoritmos Apriori, Chrono_Assoc e AssocTemp, inicia cada um dos algoritmos e grava os padrões obtidos na base de conhecimento, disponibilizando-os para análise.

É importante citar que esta operação pode ser realizada uma só vez ou então, cada vez que o especialista deseje renovar a base de conhecimento.

A análise dos padrões, por sua vez, se dá mediante uma sequência de interações entre o especialista e o modelo, descritas no diagrama de atividades da Figura 19.

A interface para análise dos padrões permite ao especialista selecionar o procedimento que deseja analisar, mostra a frequência dos procedimentos que ocorrem em conjunto com a seleção, mostra as regras de associação, padrões de sequência e os intervalos de tempo, bem como os beneficiários que estão cobertos pela regra de associação selecionada pelo especialista (Figura 20).

Adicionalmente, o especialista pode configurar como o modelo se comporta, filtrando regras nas quais o procedimento selecionado esteja no antecedente ou no consequente, exibindo somente as regras que representam implicações, em que o antecedente vem antes do consequente, exibir as regras que não possuem o procedimento selecionado ou definindo se o modelo deve exibir os dados filtrados automaticamente ou se o especialista deseja aplicar os filtros manualmente.

Nessa interface, o especialista pode verificar a frequência dos eventos encontrados na base e quais são os eventos que ocorrem em conjunto com o evento selecionado (coocorrentes). Também é possível filtrar as regras de associação e os padrões cronológicos, de maneira a melhor evidenciar os padrões obtidos. Os dados dos beneficiários envolvidos nos eventos são exibidos à medida que os padrões são selecionados, detalhando as informações que atendem a tais padrões.

O especialista pode selecionar um dos beneficiários cobertos pela regra de associação selecionada (na grade Causa e Efeito) e exibir os detalhes relativos aos procedimentos realizados por tal beneficiário (Figura 21)

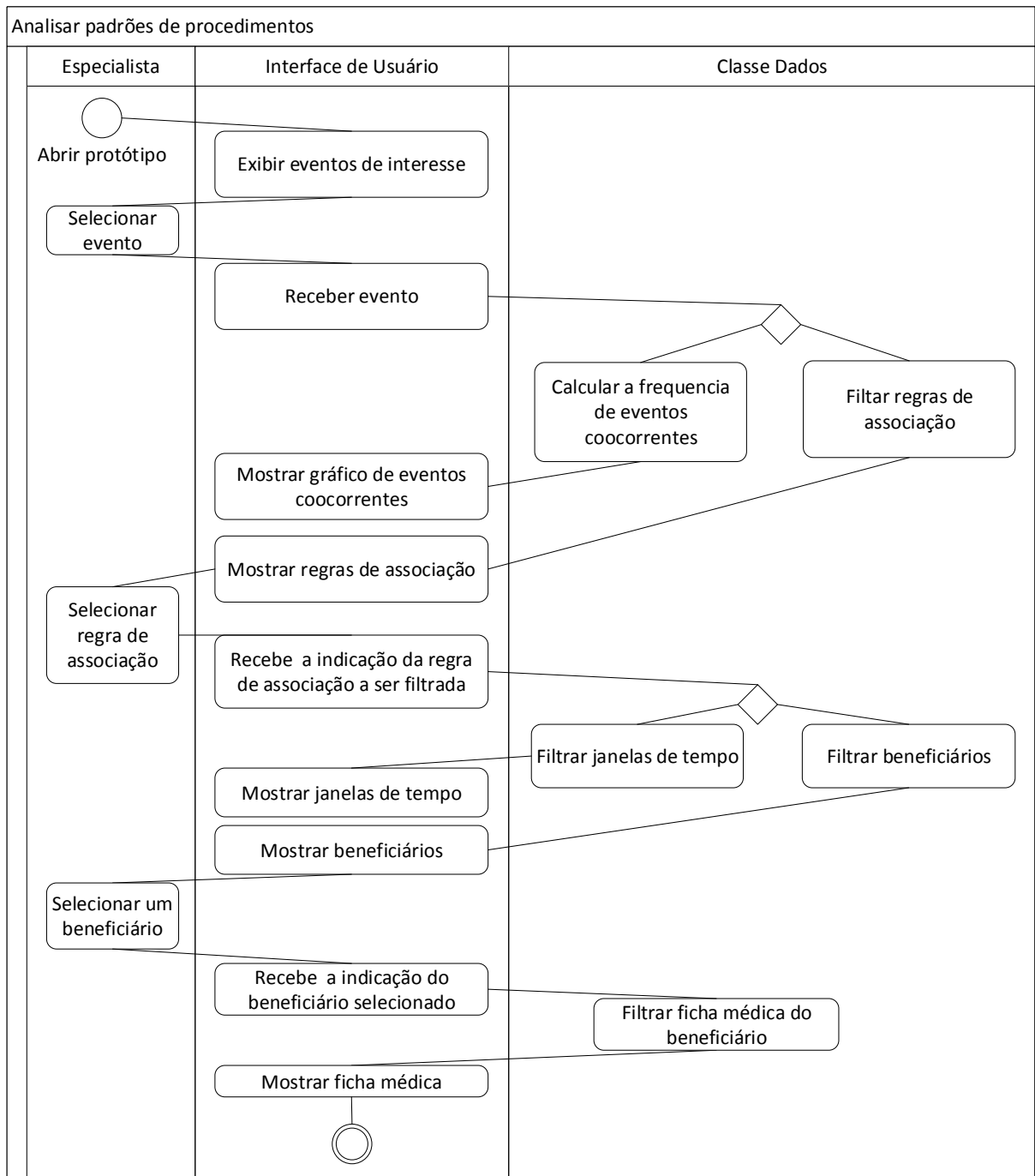


Figura 19 Diagrama de atividades - analisar padrões de procedimentos
 Fonte: O autor (2014)

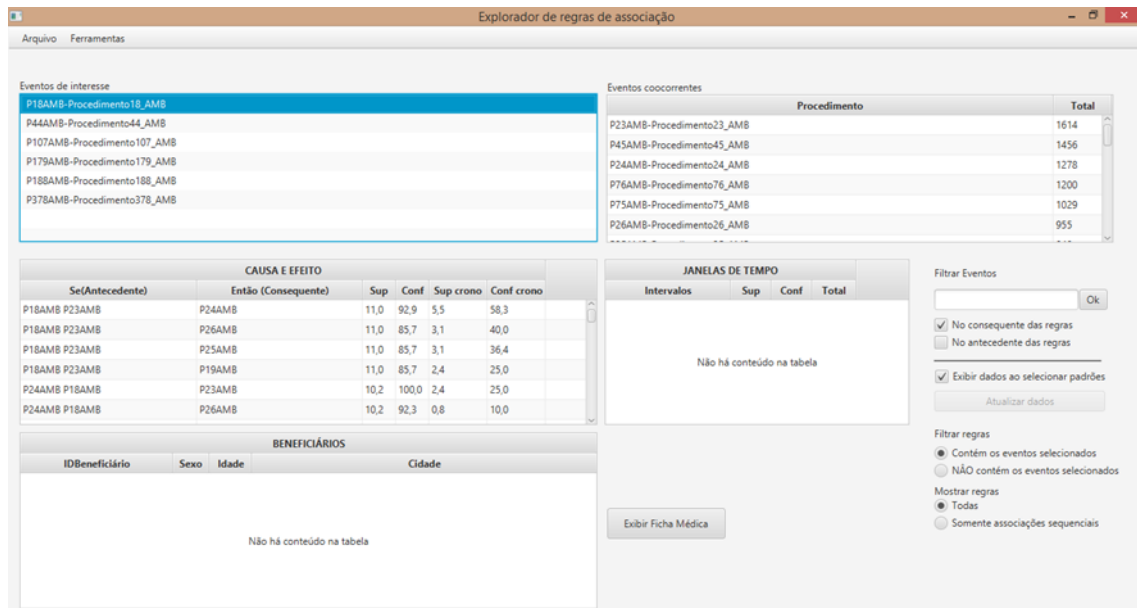


Figura 20 Interface de usuário do modelo
Fonte: O autor (2014)

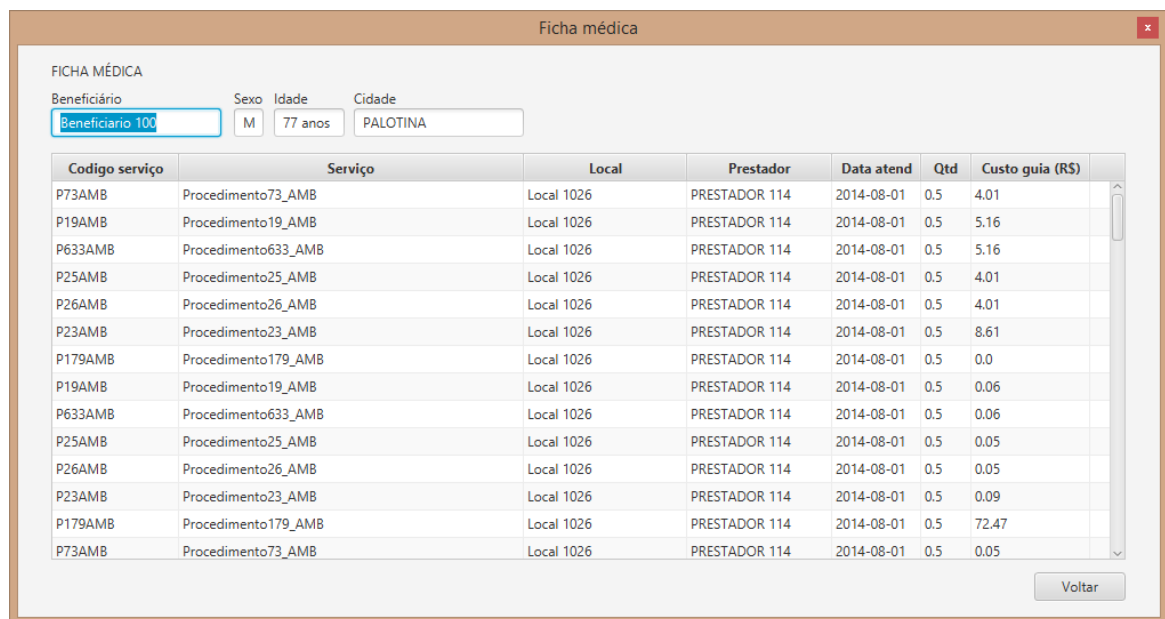


Figura 21 Ficha médica do beneficiário
Fonte: O autor (2014)

As descrições dos elementos da interface de usuário do modelo, bem com uma guia para utilização do modelo estão descritos no Apêndice B.

4.4 Padrões obtidos e base de conhecimento

Para exemplificar os padrões obtidos pelo modelo, enfatiza-se a leitura que pode ser realizada pelo especialista durante a navegação nas regras de associação, nas sequências e nos intervalos de tempo.

Na Tabela 1 são exemplificadas duas regras sequenciais, cujo suporte e confiança cronológicos são maiores que 0 (zero) e duas regras de associação não sequenciais, cujos suporte e confiança cronológicos são 0 (zero).

Tabela 1 Exemplos de regras de associação sequenciais e não sequenciais obtidas pelo modelo

	Antecedente	Consequente	Suporte	Confiança	Suporte Cronológico	Confiança Cronológica	Interpretação
1	DOSAGEM COLESTEROL e GAMA-GLUTAMIL	ÁCIDO LÁTICO	6,3%	85,7%	2,4%	42,9%	Sequencial
	LUVA CIRURGICA e GASOMETRIA	ÁCIDO LÁTICO	5,5%	85,7%	0,0%	0,0%	Não sequencial
2	ZOFRAN e DOSAGEM DE FOSFATASE ALCALINA	ANTÍGENO CARCINEGÊNICO	5,5%	85,7%	0,8%	16,7%	Sequencial
	HEPAMAX S e GAMA-GLUTAMIL TRANSFERASE	ANTÍGENO CARCINEGÊNICO	5,5%	87,5%	0,0%	0,0%	Não sequencial

Fonte: O autor (2015)

4.4.1 Interpretação dos padrões obtidos

Apresentam-se três exemplos dos padrões obtidos com o modelo proposto, de forma a sustentar a demonstração de como devem ser interpretados pelos especialistas.

Exemplo 1: Tomando a regra de associação “Se DOSAGEM DE TRANSAMINASE OXALACETICA e GASOMETRIA então ÁCIDO LÁTICO” como exemplo, interpreta-se (Tabela 2):

- Pelo suporte e confiança cronológicos: 393 (suporte cronológico = 2,4%) beneficiários realizaram DOSAGEM DE TRANSAMINASE OXALACETICA e GASOMETRIA antes e, destes, 131 (confiança cronológica = 33,3%) realizaram ACIDO LÁTICO na sequência;
- Pelos intervalos de tempo: Prevalece até 90 dias entre a ocorrência de “DOSAGEM DE TRANSAMINASE OXALACETICA e GASOMETRIA” e “ÁCIDO LÁTICO” (confiança no intervalo = 100%).

Tabela 2 Padrões descobertos para a regra de associação: DOSAGEM DE TRANSAMINASE OXALACETICA e GASOMETRIA -> ÁCIDO LÁTICO

Antecedente	Consequente	APRIORI		SEQUÊNCIA		JANELAS DE TEMPO					
		Suporte	Confiança	Suporte Cronológico	Confiança Cronológica	90 DIAS		120 DIAS		150 DIAS	
						Suporte	Confiança	Suporte	Confiança	Suporte	Confiança
DOSAGEM DE TRANSAMINASE OXALACETICA e GASOMETRIA	ÁCIDO LÁTICO	7,9%	90,0%	2,4%	33,3%	3,9%	100,0%	3,9%	100,0%	3,9%	100,0%

Fonte: O autor (2015)

Exemplo 2: Tomando a regra de associação “Se CULTURA AUTOMATIZADA e DOSAGEM DE FOSFATASE ALCALINA então ÁCIDO LÁTICO” como exemplo, interpreta-se (Tabela 3):

- De acordo com suporte e confiança cronológicos: 131 (suporte cronológico = 0,8%) beneficiários realizaram CULTURA AUTOMATIZADA e DOSAGEM DE FOSFATASE ALCALINA antes e, destes, 19 (confiança cronológica = 14,3%) realizaram ÁCIDO LÁTICO na sequência;
- Pelos intervalos de tempo: Prevalece até 150 dias entre a ocorrência de “CULTURA AUTOMATIZADA e DOSAGEM DE FOSFATASE ALCALINA” e “ÁCIDO LÁTICO” (confiança no intervalo = 80%).

Tabela 3 Padrões descobertos para a regra de associação: CULTURA AUTOMATIZADA e DOSAGEM DE FOSFATASE ALCALINA -> ÁCIDO LÁTICO

Antecedente	Consequente	APRIORI		SEQUÊNCIA		JANELAS DE TEMPO					
		Suporte	Confiança	Suporte Cronológico	Confiança Cronológica	90 DIAS		120 DIAS		150 DIAS	
						Suporte	Confiança	Suporte	Confiança	Suporte	Confiança
CULTURA AUTOMATIZADA e DOSAGEM DE FOSFATASE ALCALINA	ÁCIDO LÁTICO	6,3%	85,5%	0,8%	14,3%	3,9%	80,0%	3,9%	80,0%	3,9%	80,0%

Fonte: O autor (2015)

Exemplo 3: Tomando a regra de associação “Se MARCADORES TUMORAIS e DOSAGEM DE FOSFATASE ALCALINA então ANTÍGENO CARCINOEMBRIOGÊNICO” como exemplo, interpreta-se (Tabela 4):

- Pelo suporte e confiança cronológicos: 393 (suporte cronológico = 2,4%) beneficiários realizaram MARCADORES TUMORAIS e DOSAGEM DE FOSFATASE ALCALINA antes e, destes, 147 (confiança cronológica = 14,3%) realizaram ANTÍGENO CARCINOEMBRIOGÊNICO na sequência;

- Pelos intervalos de tempo: Prevalece até 90 dias entre a ocorrência de “MARCADORES TUMORAIS e DOSAGEM DE FOSFATASE ALCALINA” e “ANTÍGENO CARCINOEMBRIOGÊNICO” (confiança no intervalo = 100%).

Tabela 4 Padrões descobertos para a regra de associação: MARCADORES TUMORAIS e DOSAGEM DE FOSFATASE ALCALINA -> ANTÍGENO CARCINOEMBRIOGÊNICO

Antecedente	Consequente	APRIORI		SEQUÊNCIA		JANELAS DE TEMPO					
		Suporte	Confiança	Suporte Cronológico	Confiança Cronológica	90 DIAS		120 DIAS		150 DIAS	
						Suporte	Confiança	Suporte	Confiança	Suporte	Confiança
MARCADORES TUMORAIS e DOSAGEM DE FOSFATASE ALCALINA	ANTÍGENO CARCINOEMBRIOGÊNICO	7,1%	88,9%	2,4%	37,5%	3,9%	100,0%	3,9%	100,0%	3,9%	100,0%

Fonte: O autor (2015)

Exemplo 4: Tomando a regra de associação “Se DOSAGEM DE UREIA e ZOFRAN então ANTÍGENO CARCINOEMBRIOGÊNICO” como exemplo, interpreta-se (Tabela 5):

- Pelo suporte e confiança cronológicos: 131 (suporte cronológico = 0,8%) beneficiários realizaram DOSAGEM DE UREIA e ZOFRAN antes e, destes, 22 (confiança cronológica = 16,7%) realizaram ANTÍGENO CARCINOEMBRIOGÊNICO na sequência;
- Pelos intervalos de tempo: Prevalece até 90 dias entre a ocorrência de “DOSAGEM DE UREIA e ZOFRAN” e “ANTÍGENO CARCINOEMBRIOGÊNICO” (confiança no intervalo = 75%), sendo que o intervalo máximo para estas sequências é de 120 dias (confiança no intervalo de 120 dias).

Tabela 5 Padrões descobertos para a regra de associação: DOSAGEM DE UREIA e ZOFRAN -> ANTÍGENO CARCINOEMBRIOGÊNICO

Antecedente	Consequente	APRIORI		SEQUÊNCIA		JANELAS DE TEMPO					
		Suporte	Confiança	Suporte Cronológico	Confiança Cronológica	90 DIAS		120 DIAS		150 DIAS	
						Suporte	Confiança	Suporte	Confiança	Suporte	Confiança
DOSAGEM DE UREIA e ZOFRAN	ANTÍGENO CARCINOEMBRIOGÊNICO	5,5%	85,7%	0,8%	16,7%	3,1%	75,0%	3,1%	100,0%	3,1%	100,0%

Fonte: O autor (2015)

Ainda que os especialistas não tenham avaliado os padrões obtidos pelo modelo proposto, uma vez que a proposta é disponibilizar o maior número possível de padrões e permitir que o especialista interprete a utilidade e use tais padrões em

sua rotina de trabalho, a leitura dos padrões é uma questão importante para a aceitação do modelo, pois facilita a sua adoção.

4.5 Avaliação do modelo proposto

A fim de compreender como cada aspecto do modelo proposto foi avaliado pelos especialistas, a análise dos resultados foi segmentada por questão, construto e pontuação total, permitindo verificar a aceitação detalhada e total do modelo proposto e assim, avaliar sua contribuição para o uso do KDD na rotina dos especialistas.

A pontuação do modelo proposto foi determinada dividindo-se os pontos obtidos no instrumento de avaliação pelos pontos totais possíveis (10) e a avaliação utilizou os critérios definidos no método da pesquisa.

O grupo de avaliadores contou com 6 especialistas com experiência média de 8 anos em auditoria de procedimentos em operadoras de Saúde, sendo composto por 3 enfermeiros (50%), 2 analistas de sistemas (33%) e 1 médico (17%), conforme o Gráfico 2.

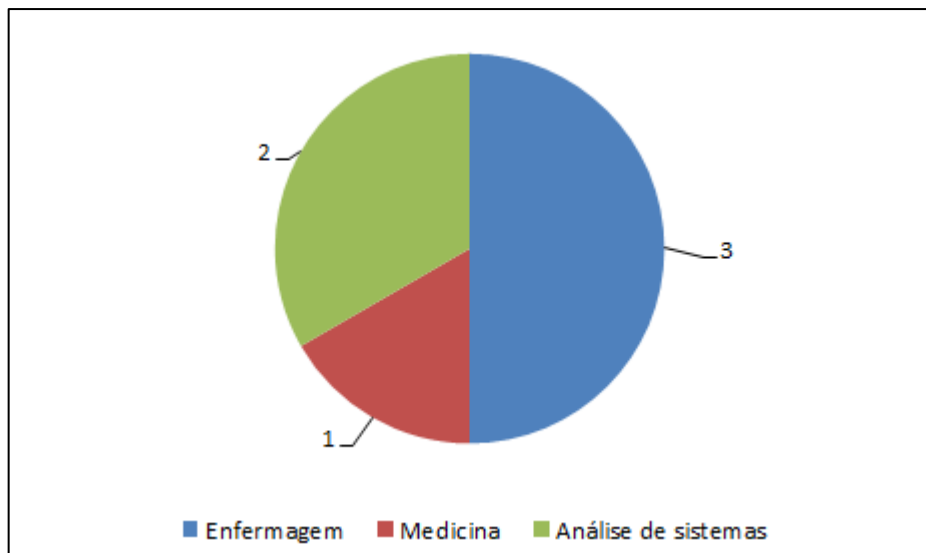


Gráfico 2 Distribuição dos especialistas quanto à formação profissional
Fonte: O autor (2015)

4.5.1 Pontuação obtida pelas questões do instrumento de avaliação

Das 10 questões consideradas no instrumento de avaliação, 3 não foram validadas, 1 foi parcialmente validada e 6 foram validadas pelos especialistas (Gráfico 3).

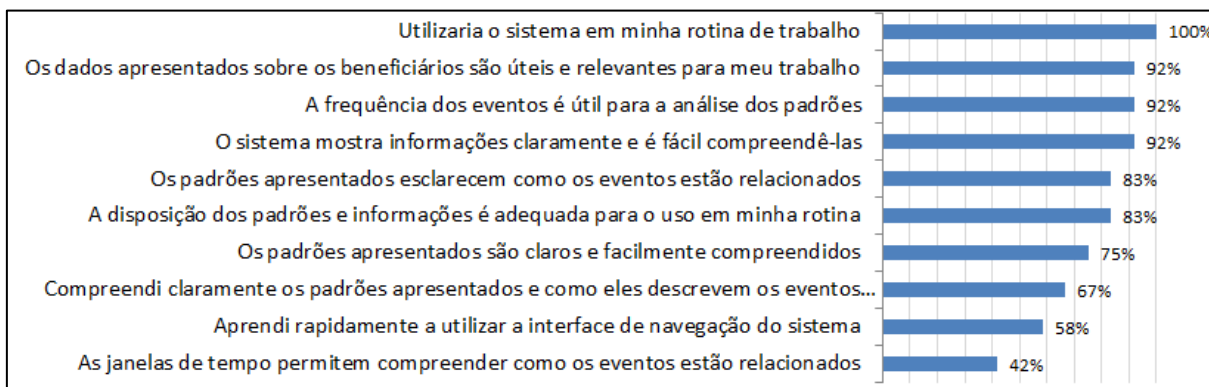


Gráfico 3 Pontuação obtida nas questões da avaliação do modelo proposto
Fonte: O autor (2015)

As questões que obtiveram as menores pontuações foram “As janelas de tempo permitem compreender como os eventos estão relacionados”, 42%; “Aprendi rapidamente a utilizar a interface de navegação do sistema”, 58% e “Compreendi claramente os padrões apresentados e como eles descrevem os eventos ocorridos”, 67%, sendo avaliadas como “não validadas” pelos especialistas.

Para essas questões, os especialistas solicitaram a inclusão dos valores nominais representados pelos padrões. Isto sugere a necessidade de relacionar as quantidades de beneficiários às quais os padrões se referem, o que facilitaria a aceitação do modelo no tocante aos quesitos relacionados com as questões que não foram validadas.

A questão “Os padrões apresentados são claros e facilmente compreendidos” obteve 75% da pontuação, sendo avaliada como parcialmente validada. Neste ponto, os especialistas também solicitaram a inclusão dos valores nominais aos quais os percentuais dos padrões se referem. Os especialistas também declararam dificuldades em interpretar os intervalos de tempo, solicitando que os intervalos não se sobreponham, mas sejam expressos como “até 90 dias”, “de 91 a 120 dias” e “de 121 a 150 dias”, demonstrando os intervalos de forma mais clara e intuitiva e, por isso, não exigindo explicações adicionais para seu

entendimento. O uso dos códigos dos procedimentos no lugar de seus títulos também foi criticado pelos especialistas.

Embora as questões “A disposição dos padrões e informações é adequada para o uso em minha rotina” e “Os padrões apresentados esclarecem como os eventos estão relacionados” tenham obtido 83% dos pontos, e por isso, consideradas validadas, o fato de estarem próximas a 80% traz a discussão quanto ao design de interação do modelo proposto para o usuário. Os especialistas mostraram insatisfação quanto à transição entre os elementos de navegação, afirmando que não ficou clara qual a sequência de seleção dos padrões para a devida navegação e exploração dos padrões. A sequência de seleção implementada está descrita na Figura 24.

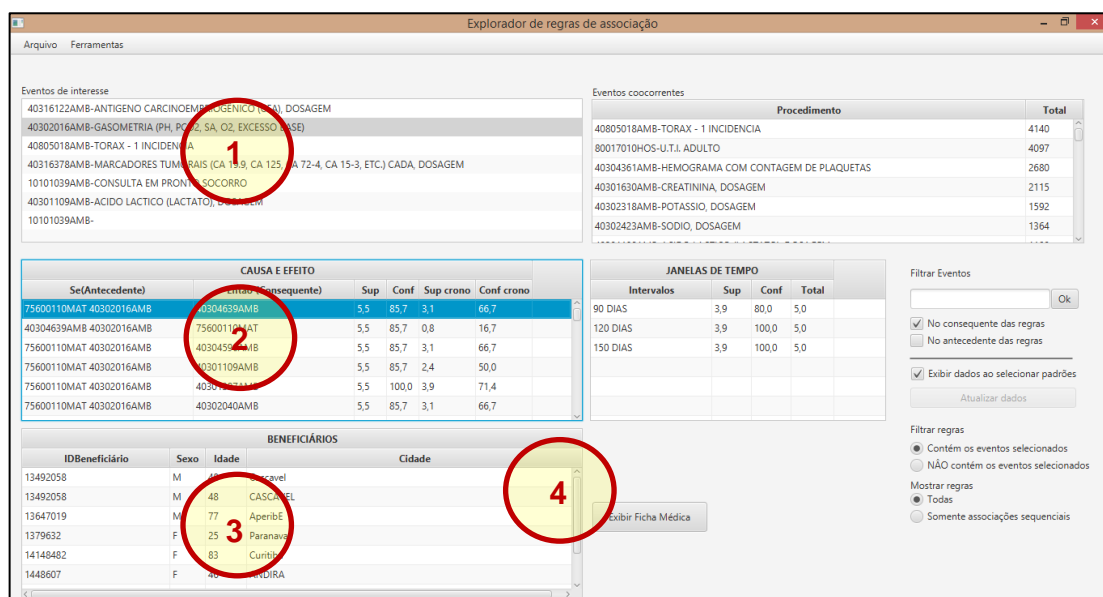


Figura 22 Sequência de navegação no modelo proposto
Fonte: O autor (2015)

Partindo destes resultados, é possível considerar que o modelo não foi capaz de atender com qualidade ao ponto de atenção “descrição de eventos ocorridos”, especialmente quando se trata da facilidade de interpretação dos padrões. O detalhamento das informações conforme solicitado pelos especialistas e a otimização da navegação, de forma que o especialista perceba melhor onde deve selecionar os elementos das interface gráfica podem compensar a baixa aceitação relativa a estes aspectos.

Por outro lado, as questões que obtiveram as melhores pontuações e, por isso, foram validadas foram: “O sistema mostra informações claramente e é fácil

compreendê-las”, 92%; “A frequência dos eventos é útil para a análise dos padrões”, 92%; “Os dados apresentados sobre os beneficiários são úteis e relevantes para meu trabalho”, 92% mostram que o modelo foi capaz de atender ao ponto de atenção “representação visual dos resultados”, pois os especialistas declararam facilidade em compreender as informações exibidas.

Já, a questão “Utilizaria o sistema em minha rotina de trabalho”, obteve 100% dos pontos e mostra que a combinação dos pontos de atenção selecionados para o modelo, juntamente com a implementação dos requisitos levantados reforçou o potencial de utilidade do modelo proposto, confirmando sua capacidade de clareza, a relevância dos padrões e a clareza das informações. Finalmente, todos os especialistas entrevistados utilizariam o modelo proposto em suas rotinas de trabalho, confirmando a hipótese de que a combinação de diferentes técnicas de pós-processamento dos padrões mediante uma estrutura de navegação simples e de fácil uso reforçam a possibilidade de adoção da Mineração de Dados na rotina da gestão em Saúde.

4.5.2 Pontuação obtida pelos construtos do instrumento de avaliação

O construto “Capacidade de descrever os eventos” obteve nota geral de 64% e não foi validado pelos especialistas. Os construtos “Satisfação do especialista quanto ao modelo” e “Apresentação dos padrões e informações” obtiveram notas gerais de 83% e 85% respectivamente, sendo considerados válidos pelos especialistas (Gráfico 4).

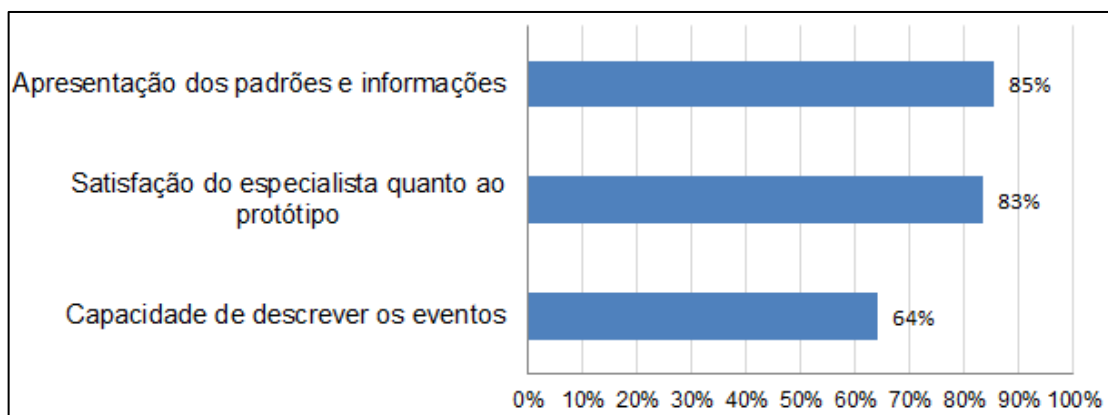


Gráfico 4 Pontuação obtida pelos construtos avaliados para o modelo proposto.
Fonte: O autor (2015)

A nota mais baixa para o construto “Capacidade de descrever os eventos” (64%) está relacionada com o fato de que duas das questões com os menores pontos estarem neste construto: “As janelas de tempo permitem compreender como os eventos estão relacionados”, com pontuação 42% e “Compreendi claramente os padrões apresentados e como eles descrevem os eventos ocorridos”, com pontuação 67%.

A avaliação dos especialistas formados em Enfermagem em relação ao construto “Capacidade de descrever os eventos” foi a mais baixa (44% dos pontos) quando comparada com a avaliação dos especialistas formados em Análise de Sistemas ou em Medicina (Gráfico 5), sugerindo que os padrões não estão descritos da maneira mais adequada para o primeiro grupo. A exibição dos valores nominais representados pelos percentuais dos suportes e confianças dos padrões, o uso dos títulos no lugar dos códigos dos eventos e uma sequência de navegação mãos amigável são alternativas a serem levadas em conta em futuros experimentos com o modelo.

Já o grupo de analistas de sistemas e o médico avaliaram o construto “Capacidade de descrever os eventos”, com 83% dos pontos ambos e, portanto, válido para esses especialistas.

Entretanto, a baixa pontuação média desse construto leva a recomendação de investir-se mais nesse quesito, pois está diretamente ligado ao ponto de atenção “Descrição de Eventos”, adotado como necessário para o modelo proposto, mas que não obteve êxito na avaliação.

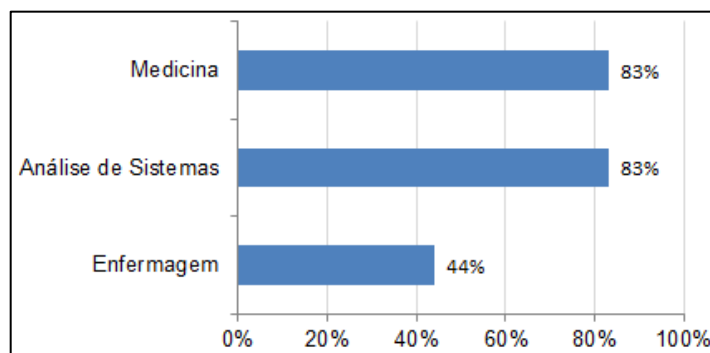


Gráfico 5 Pontuação do construto "Capacidade de descrever os eventos" em função da área de formação dos especialistas.

Fonte: O autor (2015)

O construto “Apresentação de padrões e informações” foi validado pelos enfermeiros e pelos analistas de sistemas, obtendo notas 92% e 81% respectivamente. Mas foi parcialmente validada pelo médico, com 75% (Gráfico 6).

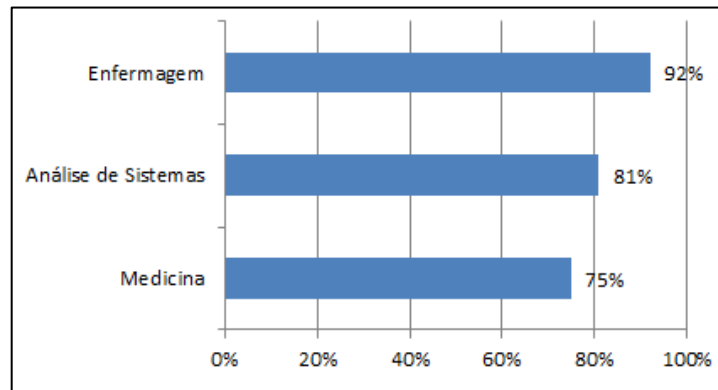


Gráfico 6 Pontuação do construto "Apresentação de padrões e informações" em função da área de formação dos especialistas.

Fonte: O autor (2015)

Um dos enfermeiros considerou que os padrões não são claros e o médico considerou que além da pouca clareza dos padrões, a disposição das informações é pouco adequada para uso em sua rotina. Mais uma vez, os especialistas solicitaram a inclusão dos valores absolutos relativos aos beneficiários cobertos pelos padrões e que os intervalos não se sobreponham, mas sejam expressos como “até 90 dias”, “de 91 a 120 dias” e “de 121 a 150 dias”.

O construto “Satisfação do especialista quanto ao modelo” foi validado pelos enfermeiros e pelos analistas de sistemas, com notas 89% e 83% respectivamente. Mas foi considerado parcialmente válido pelo médico, com nota 75%. Destaca-se a questão “Aprendi rapidamente a utilizar a interface de navegação do sistema”, que obteve 58% dos pontos, o que rebaixou a nota geral do construto e mostra que a navegação nos padrões necessitou mais atenção por parte dos especialistas e requer mais atenção quanto à sua facilidade de aprendizados e uso. (Gráfico 7).

Um dos enfermeiros, um dos analistas e o médico declaram dificuldades em aprender a utilizar a interface do modelo proposto e o médico ainda considerou que as informações apresentadas são pouco claras.

Destaca-se a questão “Aprendi rapidamente a utilizar a interface de navegação do sistema”, que obteve 58% dos pontos, o que rebaixou a nota geral do construto e mostra que a navegação nos padrões necessitou mais atenção por parte

dos especialistas e requer mais atenção quanto à sua facilidade de aprendizados e uso.

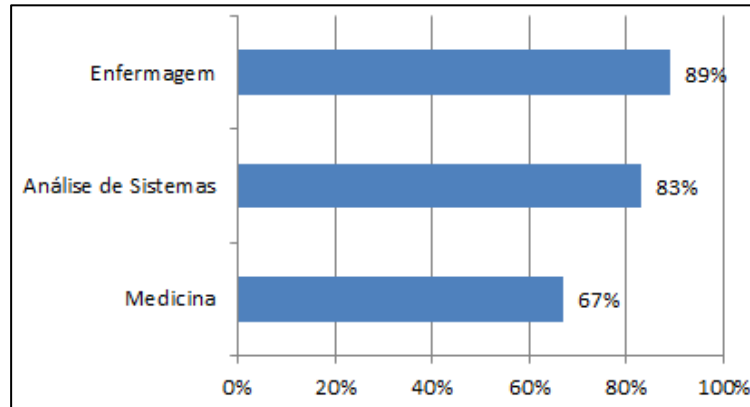


Gráfico 7 Pontuação do construto “Satisfação do especialista quanto ao modelo” em função da área de formação dos especialistas.
Fonte: O autor (2015)

A questão “O sistema mostra informações claramente e é fácil compreendê-las” alcançou 92% dos pontos, comprovando que as informações exibidas são claras e de fácil compreensão.

A pontuação geral do modelo proposto, calculado a partir da soma de todos os pontos obtidos, dividida pelo total possível de pontos, foi de 78%, sendo considerado parcialmente válido, de acordo com o IVC (WALTZ et al, 1991 apud DALLAGASSA, 2009) e com os critérios definidos no método desta pesquisa

Embora a satisfação dos especialistas quanto ao modelo proposto e a apresentação dos padrões e informação sejam consideradas válidas, a capacidade de descrever os eventos obteve a nota 64%, sendo considerada inválida no processo de avaliação. Esta nota menor está relacionada à necessidade do especialista em incluir valores nominais dos totais de beneficiários cobertos pelos padrões e de reorganizar os intervalos, eliminando a ideia de sobreposição passada por suas descrições.

5 CONSIDERAÇÕES FINAIS

As operadoras de Saúde possuem grandes quantidades de dados relativos aos atendimentos prestados a seus beneficiários e com isto, apresentam dificuldades em extrair novos conhecimentos de tais bases (DALLAGASSA, 2009), condição favorável à aplicação do KDD, por sua vocação em permitir análises mais complexas para a tomada de decisão (TRINDADE, 2005; KOBUS, 2006; SANTANA et al. 2006; LOPES, 2007; DALLAGASSA, 2009).

A motivação central para esta pesquisa está no fato de que a escassez de pesquisas sobre o uso do KDD para apoio à rotina em Saúde, frente à prevalência de aplicações de métodos estatísticos, evidencia que as análises na Área tendem a métodos quantitativos que exploram a parte visível dos dados e, portanto, outra parte – aquela que não está facilmente visível – tende a ser ignorada nas aplicações para apoio à decisão em Saúde.

Neste sentido, esta pesquisa se dedicou a discutir estratégias para que os resultados do KDD sejam adotados na rotina da gestão em Saúde, apoiando seu uso cotidiano mediante a exploração de fatores importantes para que constitua um instrumento na tomada de decisão na Saúde, propondo e avaliando um método para apoiar a eficiência na descoberta de pacientes de alto custo em operadoras de Saúde.

Partiu-se do pressuposto de que a sistematização dos pontos de atenção para uso do KDD na Saúde permite planejar como aplicações da Mineração de Dados de do pós-processamento de padrões serão construídas, quais recursos e técnicas podem ser considerados e quais tarefas da mineração devem ser implementadas conforme o problema que se deseja resolver. Assim, as aplicações do KDD na Saúde contarão com maiores chances de aceitação e uso cotidianos.

O modelo proposto manteve foco no pós-processamento dos padrões, o que se mostrou justificado, uma vez que o pós-processamento de padrões é menos frequente nos estudos pesquisados, 39% dos estudos se dedicaram a este tema, demonstrando, um espaço a ser explorado. Da mesma forma, a obtenção de regras de associação obtidas com o Apriori e pós-processadas em cadeias sequenciais, com o Chrono_Assoc, e em janelas de tempo, com o AssocTemp buscou atender ao ponto de atenção “descrever eventos ocorridos” e “fornecimento de subsídios para explicações causais”, demonstrando como as associações entre procedimentos se

comportam no decorrer do tempo, quais são as cadeias sequenciais e os intervalos relevantes na realização de procedimentos, confirmando a importância do pós-processamento de padrões para fomento do uso do KDD na Saúde.

O desenvolvimento de um sistema para suportar e facilitar a obtenção dos padrões e para auxiliar na sua exploração permitiu a aplicação dos pontos de atenção “desenvolvimento de funcionalidade específica” e “representação visual dos resultados”. Por sua vez, a interface de usuário abstraiu a complexidade relativa à filtragem das regras de associação e análise das sequências e dos intervalos de tempo, mantendo o foco dos especialistas na observação dos padrões e permitindo a rápida identificação de beneficiários cobertos pelos padrões obtidos, contribuindo para a eficiência do processo de sua identificação.

A proposta de combinação da automação da obtenção de padrões com uma base de conhecimento explorada a partir de uma interface simples e intuitiva para os usuários contribuiu para o ganho de eficiência no uso do KDD na Saúde, uma vez que o modelo proposto assume as atividades de transformação, mineração e apresentação dos padrões, cabendo à operadora de Saúde selecionar e preparar os dados e então, explorar os padrões obtidos.

O modelo apresenta pontos a melhorar no que tange à compreensão dos padrões, especialmente das janelas de tempo (AssocTemp), cuja representação ainda requer uma estratégia visual mais intuitiva, o que acabou por prejudicar o rápido aprendizado para uso.

Porém, mesmo tendo criticado a capacidade do modelo de descrever os eventos ocorridos, os especialistas se mostraram satisfeitos com a proposta, em especial com a apresentação dos padrões e das informações. Destacando-se o fato de que todos os especialistas envolvidos na avaliação declararam que utilizariam o modelo proposto em sua rotina de trabalho, o que define o sucesso da proposta, dado que o problema de pesquisa gira em torno do fomento do uso do KDD na rotina em Saúde.

Em se tratando de trabalhos futuros, pretende-se estender o modelo, em trabalhos futuros, para a adoção de mais pontos de atenção, especialmente para apoio à previsão de eventos e para a avaliação da relevância dos padrões obtidos, em uma perspectiva de ampliar as funcionalidades e possibilidades apresentadas pelo modelo proposto.

Vislumbra-se o desenvolvimento de um sistema inteligente com sensores capazes de classificar transações de acordo com a base de conhecimento e com um conjunto de atuadores, capazes de recomendar ações para potencializar eventos desejados ou ações para evitar eventos indesejados, contribuindo para o ganho de eficiência e eficácia nas decisões e, desta forma, aprofundando o uso do KDD na rotina em Saúde.

Vê-se ainda a possibilidade de pesquisas relativas à aplicação dos pontos de atenção levantados e que não foram adotados no modelo proposto, avaliando sua utilidade na Área da Saúde e, eventualmente, em diferentes áreas.

A incorporação de mais pontos de atenção é necessária, especialmente a integração de diferentes bases de dados permitirá que o modelo seja mais eficaz, por conta da capacidade de apoiar a seleção dos dados, buscando-os em diferentes fontes e auxiliando o especialista no pré-processamento dos dados como, por exemplo, na integração e redução, possibilitando a mineração de um conjunto mais amplo de dados e, conseqüentemente, a obtenção de padrões mais acurados ou representativos, e, portanto, mais úteis.

Para tanto, vislumbra-se o desenvolvimento de uma funcionalidade que automatize a seleção e preparação dos dados, conectando-se diretamente às bases de dados da operadora. Esta funcionalidade requer a elaboração da capacidade de identificar atributos mediante uma camada interna de representação que permita a apropriação de diferentes contextos, bases de dados e o atendimento a diferentes problemas relativos à decisão em Saúde.

REFERÊNCIAS

AGRAWAL Rakesh; SRIKANT Ramakrishnan. **Fast algorithms for mining association rules in large databases**. 20th International Conference on Very Large Data Bases.. p 478 – 499. 1994

BELLAZZI, Ricardo; ZUPAN, Blaz. **Predictive data mining in clinical medicine: Current issues and guidelines**.. International Journal of Medical Informatics n77 ,P 81–97. 2008

BLONBERG, Luciano Costa. **Gestão de métricas e indicadores de doenças em saúde bucal suportado por um ambiente de descoberta de conhecimento em banco de dados**. 107f Dissertação (Mestrado em Ciência da Computação). Pontifícia Universidade Católica do Rio Grande do Sul, 2010..

BORGELT, Cristian. **Apriori**-Association rule induction. 2010. Disponível em <http://www.borgelt.net/apriori.html>. Acessado em 22/07/2014

BRIN, Sergey; MOTWANI, Rajjev; ULMAN, Jeffrey D.; TSUR, Shalon. **Dynamic itemset counting and implication rules for market basket data**. SIGMOD 97 Proceedings of the 1997 ACM SIGMOD international conference on Management of data. New York P 255-264. 1997.

CANCIAN, Píndaro Secco. **Proposta de um sistema de prontuário eletrônico do paciente**. 91 f. Dissertação (Mestrado em Ciências).. CPGEI, Universidade Tecnológica Federal do Paraná, Curitiba, 2006.

CARVALHO, Deborah Ribeiro; ESCOBAR, Leandro Fabian Almeida; TSUNODA, Denise. **Pontos de Atenção para o uso da mineração de dados da saúde**. Informação & Informação, , v. 19, n. 1, p. 249-272, fev. 2014. ISSN 1981-8920. Disponível em: <<http://www.uel.br/revistas/uel/index.php/informacao/article/view/16532>>. Acesso em: 25/05/2014. doi:10.5433/1981-8920.2014v19n1p249.

CARVALHO, Deborah Ribeiro. **Árvore de Decisão / Algoritmo Genético para tratar o problema de pequenos disjuntos em classificação de Dados**. 173f. Tese (Doutorado em Ciências). Universidade Federal do Rio de Janeiro. 2005.

CARVALHO, Deborah Ribeiro; MOSER, Auristela Duarte; SILVA, Verônica Andrade; DALLAGASSA, Marcelo Rossano. **Mineração de dados aplicada à fisioterapia**. Fisioterapia e Movimento. V 25, N 3, 595 – 605. 2012.

CHEN, Ming-Syan; HAN, Jiawei; YU, Philip S. **Data Mining: An overview from database perspective**. IEEE Transactions on Knowledge and Data Engineering. V8 N6.p.866-883. 1996.

CRUZ-RAMÍREZ, Manuel; HERVÁS-MARTÍNEZ, César; GUTIÉRREZ; Pedro Antonio; PÉREZ-ORTIZ, María; BRICEÑO, Javier; DE LA MATA, Manuel. **Memetic**

Pareto differential evolutionary neural network used to solve an unbalanced liver transplantation problem. Soft Computing. V17 n2 p 275-284. 2013

DALLAGASSA, Marcelo Rosano. **Concepção de uma metodologia para identificação de beneficiário com indicativos de diabetes melitos tipo 2.** 105 f. Dissertação (Mestrado em Tecnologia e Saúde). Pontifícia Universidade Católica do Paraná. 2009.

FAYYAD, Usama; PIATETSKI-SHAPIRO, Gregory; SMYTH, Padhraic. **The KDD process for extracting useful knowledge from volumes of data.** Communications of the ACM. n11 v39 27 – 34. 1996.

FAYYAD, U.; Piatetsky-Shapiro, G.; Smyth, P.; Uthurusamy, R. **Advances in knowledge discovery and data mining.** american association for artificial intelligence. Menlo Park, CA: MIT Press. 1996

GENG, Liqiang; HAMILTON, Howard J. **Interestingness measures for data mining:** A survey. ACM Computing Surveys, V 38, N 3, 9 - 41. 2006.

GOMES, Heitor M; CARVALHO, Deborah R. **A hybrid data mining method: exploring sequential indicators over association rules.** Iberoamerican Journal of Applied Computing. v.1 n.4. Agosto, 2011.

HAN, Jiawei; KAMBER, Micheline. **Data mining:** Concepts and techniques. Elsevier. San Francisco. 2006

HUSSAIN, Farhad; LIU, Huan ; SUZUKI, Einoshin; LU, Hongjun. **Exception rule mining with a relative interestingness measure.** PAKDD-2000, LNAI v.1805, p. 86-96. 2000.

KOBUS, Luciana Schleder Gonçalves. **Identificação de usuários com doenças cardiovasculares elegíveis para programas de gerenciamento de caso – aplicação da descoberta de conhecimento em bases de dados a uma organização de saúde do tipo auto-gestão.** 146 f. Dissertação (Mestrado em Tecnologia em Saúde). Pontifícia Universidade Católica do Paraná. 2006

LIKERT, Rensis. **A Technique for the measurement of attitudes.** Achives of Psychology. University of New York. N 140 p 5 – 54. 1932.

LOPES, Lucelene. **Aprendizagem de máquina baseada na combinação de classificadores em bases de dados da área da Saúde.** 117 f. Dissertação (Mestrado em Tecnologia em Saúde). Pontifícia Universidade Católica do Paraná. Curitiba. 2007

MAGIAMELI, Paul; WEST, David; RAMPAL, Rohit. **Model selection for medical diagnosis decision support systems.** Decision Support Systems. V 36, 247-259. 2004.

MARISCAL, Gonzalo; MARBÁN, Óscar; FERNANDEZ, Covadonga. **A survey of data mining and knowledge discovery process models and methodologies**. The Knowledge Engineering Review. v25 n2, p137–166. 2010.

MEYFROIDT Geert; GÜIZA, Fabian; RAMON, Jan; BRUYNOOGHE, Maurice. **Machine learning techniques to examine large patient databases**. Best Practice & Research Clinical Anaesthesiology. n 23 p 127–143. 2009.

MILANI, Cristian. Simioni; CARVALHO, Deborah Ribeiro. **Pós-Processamento em KDD**. Revista de Engenharia e Tecnologia. v 5, p. 151-162. 2013.

MILOVIC, Boris; MILOVIC, Milan. **Prediction and decision making in health care using data mining**. International Journal of Public Health Science. V1, N2. P 69-78. 2012.

MITRA, Sushmita; ACHARYA, Tinku. **Data mining: multimedia, soft computing, and bioinformatics**. John Wiley & Sons. New Jersey. 2003

MITRA, Sushmita; PAL, Sankar; MITRA, Pabitra. **Data mining in soft computing framework: A survey**. IEEE Transactions on Neural Networks, vol. 13, pp. 3-14. 2002.

NÉRI, Eugénie Desirée Rabelo; GADÊLHA, Paulo Gean Chaves; MAIA, Sâmia Graciele; PEREIRA, Ana Graziela da Silva. **Erros de prescrição de medicamentos em um hospital brasileiro**. Revista Associação Médica Brasileira. V57 N3. 306-314. 2011.

PATEL, Vimla L.; KAUFMAN, David R.; KANNAMPALLIL, Thomas G. . **Diagnostic reasoning and decision making in the context of health information technology**. Reviews of Human Factors and Ergonomics.. n8, p149-193. 2013.

PEREIRA, Irene Mari. **Dimensionamento informatizado de profissionais de enfermagem (DIPE)**: Avaliação de um software. 148 f. Dissertação de Mestrado. Escola de Enfermagem. Universidade de São Paulo. 2011.

RAMON Jan, FIERENS Daan; GÜIZA Fabián; MEYFROIDT Geert; BLOCKEEL Hendrik; BRUYNOOGHE Maurice; VAN DEN BERGHE, Greet. **Mining data from intensive care patients**. Advanced Engineering Informatics,. n21 p243–256. 2007.

RODDICK, John F.; SPILIOPOULOU, Myra. **A bibliography of temporal, spatial and spatiotemporal data mining research**,. Proc. SIGKDD Explorations, V 1, N 1, 34-38. 1999.

SANTANA, Ádamo L. de; FRANCÊS, Carlos Renato L.; COSTA, João C. W. A.; DE TARSO, Paulo. **Observatório de saúde da Amazônia - Sistema de suporte à decisão para os gestores de saúde pública da Amazônia**. Environmental and Health World Congress. 2006.

SILBERSCHATZ, A; TUZHILIN, A. **What makes patterns interesting in knowledge discovery systems**. IEEE Transactions on Knowledge and Data Engineering, vol. 8, pp. 970-974. 1996.

SHILLABEER, Anna. Roddick, John. **Reconceptualising interestingness metrics for medical data mining**. Workshop on Health Data Mining. School of Informatics and Engineering, Flinders University of South Australia, 2005

_____. **Establishing a lineage for medical knowledge discovery**. 6th Australasian Data Mining Conference, Gold Coast, Australia. 2007

SHIMIZU, Tamio, **Decisão nas organizações**. 3ª Edição. Editora Atlas, São Paulo. 2010.

SMITH, Megan; HIGGS, Joy; ELLIS, Elizabeth. **Factors influencing clinical decision making in clinical reasoning and clinical decision making – Nature and context**. In HIGGS J; JONES M; Loftus S; Christensen N. **Clinical reasoning in the health professions 3th ed.**, New York. Butterworth-Heinemann 2008 p 89 -100.

SOKOLOSKI, Willian Felipe, CARVALHO, Deborah Ribeiro, DALLAGASSA, Marcelo Rosano. **Regra de associação temporal**. XIV Congresso Brasileiro em Informática em Saúde – CBIS, Santos, Brasil, 2014

SPRAGUE, Ralph H; WATSON, Hugh J. **Sistemas de apoio à decisão: Colocando a teoria em prática**. Rio de Janeiro. Campus. 1991

TARGINO, Maria das Graças. **Informação em saúde: Potencialidades e Limitações. Informação & informação..** V 14 n1. 2009. Disponível em <http://www.uel.br/revistas/uel/index.php/informacao/article/view/1845/2891>. Acessado em: 25/03/2015.

TAN, Pang-Nin; KUMAR, Vipin; SRIVASTAVA, Jaydeep. **Selecting the right interestingness measure for association patterns**. Proceedings of the 8th International Conference on Knowledge Discovery and Data Mining (KDD 2002). Edmonton, Canada. 32–41. 2002.

TRINDADE, Carla Machado da. **Identificação do comportamento das hepatites virais a partir da exploração de bases de dados de saúde pública**. 141 f. Dissertação (Mestrado em Tecnologia em Saúde). Pontifícia Universidade Católica do Paraná. 2005.

TURBAN, Efraim ; ARONSON, Jay E. **Decision support systems and intelligent systems** Upper Sadle River. Prentice Hall, 2001

TURBAN, Efraim; SHARDA, Ramesh; ARONSON, Jay E.; KING, David. **Business Intelligence: Um enfoque gerencial para a inteligência do Mercado**. Porto Alegre. Bookmann. 2009

APÊNDICES

APÊNDICE A

Artigo publicado no periódico Informação e Informação, sob DOI: 10.5433/1981-8920.2014v19n1p249

PONTOS DE ATENÇÃO PARA O USO DA MINERAÇÃO DE DADOS NA SAÚDE **Attention Points in Data Mining For Decision Support in Health** **Puntos de Atención En Minería De Datos De Apoyo A Las Decisiones En Salud**

Deborah Ribeiro Carvalho - ribeiro.carvalho@pucpr.br
 Doutora em Computação de Alto Desempenho pela Universidade Federal Do Rio Janeiro (UFRJ). Professor da Pontifícia Universidade Católica do Paraná (PUCPR).

Leandro Fabian Almeida Escobar - l.escobar72@gmail.com
 Mestrando do Programa de Pós-Graduação em Ciência e Gestão da Informação da Universidade Federal do Paraná (UFPR). Professor da Universidade Positivo (UP).

Denise Tsunoda - dtsunoda@gmail.com
 Doutora em Engenharia Elétrica e Informática Industrial pela Universidade Tecnológica Federal do Paraná (UFPR). Professora da UFPR.

RESUMO

Introdução: A Mineração de Dados representa uma alternativa para apoiar a decisão, mas não constitui prática regular nas atividades da Saúde.

Objetivo: Revisar a literatura, identificando pontos de atenção na utilização da Mineração de Dados na área da Saúde.

Metodologia: Os pontos de atenção foram obtidos mediante a análise de publicações em periódicos, sistematizados em quadros de referência cruzada e foram descobertas de regras associação e respectivas exceções.

Resultados: Foram identificados 14 pontos, formando um conjunto de dados dos quais obtiveram-se 345 regras gerais e respectivas exceções. Os pontos mais citados são “previsão de eventos” e “auxílio ao planejamento” e os menos citados são “apropriação de protocolo específico” e “detecção de padrões em tempo real”. Percebe-se associação entre o “auxílio ao planejamento” e a “não detecção de padrões em tempo real”. Exceção ocorre, quando o “auxílio ao planejamento” é citado juntamente com “explicações causais”, associando-se à “detecção de padrões em tempo real”.

Conclusões: Os pontos de atenção indicam critérios para a adoção da Mineração de Dados na Saúde. Destaca-se “explicações causais” que, citado em 8 artigos, determina exceções nas associações entre os demais pontos, indicando que a seleção da tarefa de mineração passa pela adequação às expectativas dos usuários.

Palavras chave: Mineração de Dados em saúde. Pontos de atenção.

RESUMEN

Introducción: La Extracción de Datos representa una alternativa para apoyar la decisión, pero no constituye práctica regular en las actividades de la Salud.

Objetivo: Revisar la literatura, identificando puntos de atención en la utilización de la Extracción de Datos en el área de la Salud.

Metodología: Los puntos de atención fueron obtenidos mediante el análisis de publicaciones en periódicos, sistematizados en cuadros de referencia cruzada y fueron descubiertas de reglas asociación y respectivas excepciones.

Resultados: Fueron identificados 14 puntos, formando un conjunto de datos de los que se obtuvieron 345 reglas generales y las respectivas excepciones. Los puntos más citados son “previsión de eventos” y “auxilio a la planificación” y los menos citados son “apropiación de protocolo específico” y “detección de estándares en tiempo real”. Se percibe asociación entre el “auxilio a la planificación” y la “no detección de estándares en tiempo real”. Excepción ocurre, cuando el “auxilio a la planificación” es citado juntamente con “explicaciones causales”, asociándose a la “detección de estándares en tiempo real”.

Conclusiones: Los puntos de atención indican criterios para la adopción de la Extracción de Datos en la Salud. Se destacan “explicaciones causales” que, citado en 8 artículos, determina excepciones en las asociaciones entre los demás puntos, indicando que la selección de la tarea de extracción pasa por la adecuación a las expectativas de los usuarios.

Palabras clave: Extracción de Datos en la salud. Puntos de atención

ABSTRACT

Introduction: Data Mining is an alternative to support the decision making process, but it is not a regular practice in health care activities.

Objective: Review of the literature to identify points that deserve attention in the use of Data Mining in Health Care.

Methodology: These points were obtained by analysis of publications in journals, and organized in cross-reference tables. Association rules and exceptions were discovered.

Results: Fourteen points were identified, forming a set of data from which 345 general rules and their respective exceptions were obtained. The points most often cited were “prediction of events” and “support to planning” and the least cited points were “appropriation of specific protocol” and “real-time pattern recognition”. An association was observed between “support to planning” and “non real-time recognition of patterns”. An exception occurs when “support to planning” is cited along with “causal explanations”, being associated to “real-time pattern recognition”.

Conclusions: The points of attention indicate criteria for the adoption of Data Mining in Health Care. The point “causal explanations” deserves mention. It was cited in 8 articles and determines exceptions in the associations among other points, indicating that the selection of the mining task involves the adaptation to users' expectations.

Keywords: Data Mining in Health Care. Points of attention.

1.Introdução

A Mineração de Dados constitui uma alternativa para processar grandes volumes de dados dos Sistemas de Informação em Saúde, dada a sua capacidade de descobrir padrões úteis, novos e surpreendentes, possibilitando o apoio em análises complexas sobre dados clínicos.

O KDD (*Knowledge Discovery in Databases*), por sua vez, é um processo geral para a conversão de dados brutos em informações úteis (FAYYAD *et al*, 1996) e tipicamente constituído pelas etapas de pré-processamento, que envolve a de seleção, limpeza e preparação dos dados; processamento, que trata da descoberta de padrões mediante algoritmos de Mineração; pós-processamento, que refina os resultados obtidos durante o processamento, seja compondo novos padrões ou avaliando seu interesse, e interpretação dos padrões extraídos, culminando na obtenção de conhecimentos antes ocultos.

Vários são os desafios para a utilização da Mineração de Dados que se inicia com a preparação das bases de dados, seleção dos algoritmos que melhor se adequam ao problema proposto e por fim análise dos resultados, ou seja, dos padrões descobertos. Por isso mesmo, apesar de todo desenvolvimento já realizado, esta área continua sendo objeto de pesquisas por novas soluções que se aproximem do real interesse dos potenciais usuários. As razões pela constante pesquisa se devem ao fato de que, apesar da existência de várias experimentações quanto ao uso de Mineração de Dados, ainda é baixa a sua adoção nos processos de tomada de decisão diários (MARISCAL *et al*. 2010), por conta da pouca familiaridade dos especialistas com a metodologia, vantagens e dificuldades (MEYFROIT *et al* 2009), a prevalência de estudos estatísticos com objetivo de revelar relações lineares simples entre os fatores de saúde (CRUZ-RAMIREZ, *et al* 2012) ou, ainda, a exploração comparativa de técnicas de Mineração de Dados, deixando de lado etapas relativas à interpretação e avaliação de resultados (BLOMBERG, 2010).

Portanto, para que a Mineração de Dados seja efetivamente incorporada ao processo de decisão em Saúde, faz-se necessário identificar quais os pontos de atenção durante o processo de descoberta de conhecimento nas bases de dados de forma a cumprir com os critérios de aceitação determinados pelos especialistas. Parte-se da hipótese, então, de que existem elementos e situações que requerem tratamento criterioso por parte do pesquisador para que a Mineração de Dados seja adotada como ferramenta efetiva de apoio à decisão em Saúde.

Este artigo se propõe, desta forma, a pesquisar e discutir possíveis pontos de atenção que venham a facilitar uma utilização mais intensiva da Mineração de Dados na Saúde. A pesquisa se baseia em relatos sobre experiências de aplicação da Mineração de Dados na Saúde, buscando identificar questões que facilitem ou não a sua adoção.

2. Encaminhamentos metodológicos

Uma pesquisa aplicada, qualitativa, exploratória e bibliográfica foi conduzida nas bases de Periódicos Capes, Scielo, PubMed e Domínio Público. Para a pesquisa foram utilizados os seguintes descritores: Data Mining, Mineração de Dados, KDD e *Acceptance*, KDD e *Obstacle*, *Data Mining* e *Acceptance* e *Health*.

Após a leitura e análise dos textos que atenderam ao descritor foram selecionados aqueles cujo tema relata a aplicação da Mineração de Dados e a identificação de estratégias para a efetividade no uso dos padrões descobertos para

a área da Saúde. Os critérios de inclusão adotados foram: data de publicação posterior a 2005; apresentar a aplicação do KDD, bem como uma avaliação da efetividade do uso da Mineração de Dados discutindo benefícios e/ou dificuldades.

Foram identificados os aspectos citados pelos autores como importantes, tanto para a realização dos experimentos, como também para a devida incorporação na rotina dos especialistas/gestores. Estes aspectos foram agrupados por similaridade, conforme a natureza do problema, dificuldade ou solução caracterizando pontos de atenção.

Os dados coletados foram, sistematizados em uma tabela, sendo os pontos de atenção representados nas linhas os artigos contendo os relatos nas colunas. Cada célula (cruzamento das linhas e colunas) foi preenchida com S(im) ou N(ao), indicando presença ou ausência. O conjunto de dados oriundo desta tabela submetido à tarefa de Descoberta de Regras de Associação.

Para a Mineração de Dados foi utilizado o algoritmo Apriori do ambiente WEKA (WITTEN *et al*, 2011), que descobre regras de associação do tipo $X \rightarrow Y$ (se X então Y). Sendo X e Y representando itens de dados do conjunto de entrada. Para cada regra de associação descoberta são determinados o suporte da regra, indicando a ocorrência de X e Y em relação à base de dados e a confiança da regra, que indica a quantidade de registros que possuem X e também possuem Y. (AGRAWAL e SRIKANT, 1994).

As regras descobertas foram pós-processadas, para a identificação de possíveis exceções, pois tendem a ser mais interessantes que a respectiva regra geral (HUSSAIN *et al*, 2000).

Regra Geral: $X \rightarrow Y$ (se X então Y)
 Regra de Exceção: $X, A \rightarrow \neg Y$ (se X e A então não Y)

Para a descoberta das regras de exceção foi utilizado o DRE – Descobre Regras de Exceção (MILANI E CARVALHO, 2013).

3. Resultados

Foram encontradas 660 publicações (Tabela 1), das quais 625 publicações foram excluídas por não estarem de acordo com os critérios de inclusão. Os 35 artigos remanescentes foram analisados, sendo novamente excluídos 17, por se limitarem a apresentar e explicar o KDD, não discutirem estratégias ou critérios para efetividade no uso da Mineração de Dados ou por não terem foco na Área da Saúde.

Resultando assim 18 publicações que estavam de acordo com os critérios de elegibilidade (Tabela 1).

Tabela 1 Número de Artigos Encontrados e Analisados, segundo Bases de Periódicos

Base Consultada	Artigos encontrados	Artigos analisados
Domínio Público	52	4
Periódicos da Capes	398	10
Pubmed	15	0
Scielo	195	4
TOTAL	660	18

Analisando as 18 publicações percebe-se que 83% relatam experimentos envolvendo a área clínica e apenas 27% envolvendo gestão da saúde. Sobre as três grandes etapas do KDD, todos os autores relatam Mineração de Dados, 78% também discutem a etapa de Pré-Processamento e apenas 39% o Pós-Processamento.

. A relação de pontos de atenção obtida a partir da leitura e análise das 18 publicações é:

- **Apropriação de protocolo específico:** atenção aos processos específicos de diagnóstico ou tratamento de acordo com indicação formal da especialidade médica a partir de protocolos ou diretrizes amplamente aceitos pela comunidade especializada.
- **Avaliação da qualidade dos dados:** trata da atenção e crítica à consistência, disponibilidade e tratamento dos dados. Este ponto de atenção está diretamente relacionado às etapas de seleção e preparação dos dados. Ainda que a qualidade dos dados faça do problema do desenvolvimento de Sistemas de Informação em geral, sua presença se justifica pela necessidade de se melhorar a qualidade da coleta de dados, mesmo ocorrendo em sistemas específicos ou ocasião diferente do momento da tomada de decisão em Saúde.
- **Integração de diferentes bases de dados:** Enriquecimento da base de dados a partir de diferentes fontes, sejam elas distribuídas, em formatos diversos ou ainda pertencentes à organizações diferentes da organização objeto do estudo. Um exemplo é a necessidade de complementar dados sobre atendimentos com dados sobre o perfil demográfica das regiões amostras do estudo.
- **Desenvolvimento de funcionalidade específica:** Construção de software específico para uso ou exploração dos modelos gerados. Desenvolvimento de sistema ou interface de usuário que suporte o processo de tomada de decisão tanto para a visualização dos dados quanto para a navegação nos padrões encontrados.
- **Obediência a processo de trabalho específico:** O sistema de coleta de dados e de apoio à decisão reproduz ou segue um conjunto de procedimentos encadeados logicamente e definidos como modo de trabalho dentro da organização em saúde. Trata-se da reprodução de processo de trabalho particular.
- **Utilização de modelo de informações específicas:** Adoção de conjunto de informações específicas e previamente definidas seja por protocolos ou pela própria organização Saúde. Trata-se da determinação de quais informações são relevantes sem que sejam indicadas as sequências lógicas dos procedimentos ou ainda a ordem das atividades de atendimento.
- **Avaliação (subjetiva ou objetiva) da relevância dos padrões encontrados:** Avaliar os padrões extraídos determinando o quanto são relevantes ou interessantes para o especialista em saúde.
- **Combinação de diferentes tipos de tarefas de Mineração de Dados:** Uso de mais de uma tarefa de Mineração de Dados para a realização do experimento ou identificação de padrões.
- **Deteção de padrões em tempo real:** Extrair e apresentar padrões simultaneamente à ocorrência dos eventos que são fontes dos dados.

- **Auxílio ao planejamento em saúde:** Apoiar ações com base nos modelos extraídos pela Mineração de Dados.
- **Representação visual dos resultados:** Criação de recursos que facilitem a compreensão dos modelos extraídos mediante a visualização dos resultados. Trata da necessidade de comunicação amigável dos padrões encontrados para os especialistas em Saúde envolvido nos estudos.
- **Descrição de eventos ocorridos numa determinada população:** Descrever padrões aos quais uma determinada amostra está sujeita, explicando as relações encontradas entre os dados minerados.
- **Previsão da ocorrência de eventos numa determinada população:** Antecipar eventos de acordo com os padrões identificados nas bases de dados.
- **Fornecimento de subsídios para explicações causais:** Demonstração das cadeias de eventos, suas causas e consequências.

4. Síntese dos estudos analisados

Trindade (2005) aplicou o KDD para a identificação de padrões de comportamento das Hepatites Virais nas bases de dados do SINAN (Sistema de Informação de Agravos e Notificações) do Sistema Único de Saúde – Governo Federal do Brasil, objetivando subsidiar ações de controle e prevenção da doença.

Foi aplicado o algoritmo C4.5 (QUINLAN, 1993) na descoberta de padrões e o C4.5Rules para pós-processar os padrões que caracterizam as Hepatites Virais.

Recursos visuais, como gráficos e mapas, foram utilizados para facilitar o entendimento dos padrões encontrados para o especialista.

A baixa qualidade dos dados foi um fator determinante. Dos 134 atributos selecionados, 65 (48%) puderam ser utilizados. Dentre os problemas destacados, está a ausência do dado, sugerindo falhas durante coleta dos dados.

West *et al* (2005) comparam a capacidade de generalização de 24 diferentes algoritmos de aprendizado supervisionado com duas bases de dados públicas, com o objetivo de testar a acurácia de seus resultados na previsão da ocorrência do câncer de mama e de identificar suas principais causas. Em todas as execuções, é utilizada validação cruzada para os resultados mediante 100 fragmentos das bases de dados.

Segundo suas conclusões, os modelos cujo treinamento e testes foram mais instáveis (pela diversificação dos conjuntos de dados) atingem melhores níveis de acurácia, reforçando a importância da qualidade dos dados de entrada dos experimentos.

Ainda, os autores afirmam que modelos resultantes de combinações de vários algoritmos obtêm maior acurácia na previsão de eventos, em detrimento da seleção de modelos únicos.

Helma e Kazius (2005) organizam um estudo para construção e validação de um modelo preditivo para processos bioquímicos e toxicidade para uso em estudos clínicos, epidemiológicos ou sobre reações adversas, com base nas orientações da OECD (Organization for Economic Cooperation and Development).

Embora o estudo utilize uma base de dados experimentais, o texto aborda os impactos da baixa qualidade de dados, destacando a duplicidade de registros, a ausência de dados e o erro humano de preenchimento. Os autores citam o nível de agregação dos dados, por interferir nos resultados da Mineração, uma vez que os

atributos utilizados podem ser generalizados ou específicos, alterando a definição dos conjuntos de dados por conta da duplicação de valores proporcionada.

A atenção a protocolos regulatórios é evidente, por conta da necessidade em validar os padrões mediante definições formais tanto para os dados quanto para os modelos preditivos, mediante o conhecimento corrente da área específica.

Kobus (2006) aplicou o algoritmo Apriori (AGRAWAL e SRIKANT, 1994) em uma base de pacientes com mais de 40 anos de idade e com ao menos um registro relativo a doenças cardiovasculares entre 2002 e 2004, a fim de identificar associações que indiquem padrões de risco e, portanto, possíveis tratamentos de alta complexidade e alto custo.

Os resultados da Mineração foram classificados pelos especialistas quanto a sua relevância, irrelevância ou insignificância. Destaque para o fato de que regras cuja confiança é de 50%, denotando aleatoriedade (o que as classificaria como irrelevantes sob uma perspectiva objetiva) foram dadas como relevantes pelos especialistas, por conta da combinação de elementos que demonstram.

A autora discorre sobre a qualidade dos dados nos Sistemas de Informação de Saúde, o que pode inviabilizar a utilização da Mineração de Dados com recurso cotidiano na Saúde, por conta do tempo requerido na preparação e limpeza.

Steiner *et al.* (2006) aplicaram a Mineração de Dados em uma base de 118 pacientes apresentando quadro de Icterícia por câncer ou cálculo biliares. Seus experimentos se basearam no uso de técnicas de análise exploratória dos dados, realizando o pré-processamento para diferentes tarefas de Mineração de Dados visando obter aquela que discriminasse padrões com a máxima acurácia, ampliando o respaldo à tomada de decisão dos especialistas médicos quanto aos seus diagnósticos.

Destacam que a geração de árvores de decisão, obstante ser a técnica menos precisa, é aquela que conquistou maior eficiência junto aos especialistas, dada a facilidade em comunicar os padrões encontrados, deixando claro quais são os atributos discriminadores e seus respectivos pontos de corte quando comparados com os protocolos específicos de diagnóstico e tratamento e mediante julgamento do especialista. Característica interessante, segundo o estudo, para a adoção sistemática da Mineração de Dados no diagnóstico e tratamento da Icterícia.

Ramon *et al.* (2007) utilizam Mineração de Dados com o objetivo de primeiramente descobrir problemas em 1.548 pacientes de Unidades de Tratamento Intensivo (UTI) e, desta forma, encontrar padrões que apoiem o especialista intensivista no diagnóstico e no tratamento a ser aplicado em tempo real. Os autores buscam, ainda, prever riscos aos quais os pacientes estejam expostos, utilizando algoritmos de Árvore de Decisão, First Random Forests e Redes Bayesianas, todos avaliados mediante validação cruzada.

Os experimentos preveem o tempo de permanência e as chances de sobrevivência, bem como o desenvolvimento de estados de risco de vida, mediante a apropriação de critérios clínicos de avaliação de severidade APACHE II (Acute Physiology and Chronic Health Evaluation II), com base na resposta inflamatória sistêmica ou insuficiência renal, acompanhados de modelos específicos de informação clínica.

Os autores relatam dificuldades relativas a ruído nos dados, tamanho das bases de dados, existência de diferentes bases de dados a serem integradas e características individuais dos pacientes, que apresentam diferentes respostas a medicamentos ou diferentes parâmetros de saúde no momento em que ingressam na UTI.

Stein Junior (2008) aplicou o KDD em uma base de dados sobre riscos em micro regiões na cidade de Curitiba, obtida mediante um questionário de entrevista com especialistas em Saúde Coletiva, do qual foram retirados os atributos que foram processados com o algoritmo J48, com objetivo de especificar um sistema de informações para monitoramento de condições em micro áreas urbanas e apoio ao planejamento de ações para melhoria da qualidade de vida da população.

O autor declara que o fato de ter adotado atributos determinados pelos especialistas entrevistados pode ter causado uma ausência de padrões novos. Reforçando a necessidade de gerar associações que surpreendam o tomador de decisão sem a contaminação dos repertórios do especialista.

O alinhamento das regras extraídas com um protocolo prévio e relativo ao domínio em questão apoia a tomada de decisão. Por exemplo, regras que associavam condições de precariedade da urbanização de uma micro área com baixo risco para a população foram descartadas pelos especialistas, justamente porque o critério de avaliação adotado determina que uma área somente pode ser considerada de baixo risco se não apresentar precariedades.

Bodini Junior (2009) utiliza a base de dados do Sistema de Movimento de Autorização de Internação Hospitalar do Sistema Único de Saúde (AIH/SUS) na busca de *outliers*, por sua dissimilitude ou inconsistência e para uso em auditorias. Destaca que os dados, além de volumosos, são multidimensionais, cujos atributos não podem ser considerados separadamente. O autor coloca a Mineração de Dados como uma técnica indicada para a seleção do escopo de auditoria, uma vez que considera a ótica multidimensional dos dados.

Um aplicativo foi implementado para a execução da tarefa de agrupamento nebuloso. Para a Classificação de Classe Única mediante SVM – One Class, o autor se aproveitou do aplicativo LIBSVM, desenvolvido por CHANG, C.; LIN, C. (2001)

Por fim, a tarefa de Mineração é facilitada por um aplicativo com interface gráfica de usuário, que recebe os arquivos de dados, transforma-os e submete-os aos aplicativos responsáveis pelos algoritmos de Mineração. Este aplicativo aplica uma função de similaridade média aos registros normalizados da base de dados, classificando-os a seleção de uma linha de corte que evidencie os registros anômalos.

Dallagassa (2009) aplicou Árvores de Decisão C4.5 a uma base de dados de beneficiários de uma operadora privada de saúde com o objetivo de identificar relações que indicassem a possibilidade de desenvolvimento de Diabetes Melitus do tipo 2, tendo como base o protocolo de diagnóstico deste mal.

As regras encontradas foram classificadas quanto ao seu interesse utilizando as medidas objetivas de interesse Taxa de Acerto e Cobertura, agrupando-as em três classes.

Em seguida, especialistas em Saúde as assinalaram quanto a concordando, concordando parcialmente ou discordando da regra. Atestando, assim, se as regras confirmam ou não seus conhecimentos prévios.

Uma interface para navegação e consulta aos padrões encontrados foi desenvolvida pelo autor e, em conclusão, todas as etapas realizadas são organizadas de forma a propor um processo para desenvolvimento de soluções baseado no KDD.

Kuretzki (2009) integra a Mineração de Dados ao módulo Analisador, do SINPE® (Sistema Integrado de Protocolos Eletrônicos). São construídas duas funcionalidades para a seleção de dados a partir dos protocolos presentes no SINPE, executar a sua classificação mediante o algoritmo ID3 (QUINLAN, 1986) e

executar o algoritmo Apriori (AGRAWAL e SRIKANT, 1994) sobre tal base, de forma a integrar os resultados obtidos para o apoio à decisão em saúde.

Conforme sua pesquisa, 53,85% (Cinquenta e três vírgula oitenta e cinco por cento) dos entrevistados consideram excelente a possibilidade de possuir um ferramenta de Mineração de Dados no sistema SINPE, destacando a independência e facilidade de uso e leitura dos padrões por parte dos usuários. A satisfação em relação ao aplicativo é de 80% (oitenta por cento), dos quais, 40% declararam excelência do software.

Meyfroidt *et al*(2009) utilizam diferentes algoritmos para avaliação de resultados em tempo real dos tratamentos impostos a pacientes internados em Unidades de Tratamento Intensivo. Apresentando os benefícios de diferentes tarefas de aprendizado supervisionado, a fim de prever o estado geral, exposição a riscos ou tempo de permanência dos pacientes, bem como o auxílio ao planejamento de tratamentos. Os resultados de Árvores de Decisão, Florestas Aleatórias (*Random Forests*), Redes Neurais, Redes Bayesianas, *Support Vector Machines* e Processo Gaussiano são comparados com protocolos de aferição clínica quanto às suas porcentagens de acerto ou erro nas previsões e, em todos os casos, vantagens na geração de modelos preditivos são apresentadas e discutidas.

Já que nenhuma das técnicas se mostrou superior às demais, os autores recomendam a aplicação de múltiplos algoritmos sempre que possível, na busca de resultados mais acurados.

Os estudos citam a confidencialidade, a quantidade de dados, a necessidade de integração de diferentes bases de dados, sua organização e qualidade como barreiras no uso de bancos de dados clínicos para pesquisa e desenvolvimento de soluções. Ainda, afirmam que a representação simplificada dos resultados amplia o entendimento por parte do especialista, contribuindo para a adoção cotidiana dos resultados da Mineração de Dados.

Ayed *et al* (2010) constroem um sistema de apoio à decisão aplicando o KDD em Unidades de Tratamento Intensivo com o objetivo de apoio aos intensivistas no entendimento das relações causais, prevenção e previsão de infecções hospitalares.

Os autores destacam que o sucesso de um Sistema de Suporte à Decisão baseado no KDD depende da devida análise das necessidades do tomador de decisões, determinação das atividades preparatórias para apoio à decisão, manipulação de dados relevantes e visualização adequada dos resultados. Evidenciando a importância de se correlacionar os dados com as atividades humanas envolvidas na tomada de decisão.

Pregam, portanto, uma abordagem que coloque o usuário no centro do processo de desenvolvimento dos sistemas de apoio à decisão baseados no KDD, planejando e avaliando tanto as atividades quanto os resultados obtidos desde o início e até o final do processo de desenvolvimento, pois um sistema de apoio à decisão difícil de utilizar é geralmente abandonado pelo usuário.

Vianna *et al* (2010) aplicam tarefas de Classificação para descoberta das características da mortalidade infantil, formas de evitá-la e respectivas causas, em bases de dados de saúde pública no estado do Paraná.

Todo o procedimento foi baseado nos protocolos de saúde infantil e apoiados por especialistas em Saúde para validação dos padrões encontrados.

A integração de dados oriundos de três diferentes bases de dados foi fundamental para realização das tarefas de Mineração de Dados. Em paralelo, a

qualidade das informações também é considerada, uma vez que dados inconsistentes foram encontrados e corrigidos.

Lee *et al* (2011) tratam a incerteza sobre a dosagem e a frequência do tratamento com prótons para o câncer de próstata mediante a Mineração de Dados, e desenvolvem um aplicativo para planejamento de tratamentos em tempo de atendimento, integração os dados dos pacientes, oriundos de diferentes fontes e a validação dos padrões encontrados conforme os protocolos de conformidade do tratamento com prótons.

O autor cita distribuição dos dados dos pacientes em diferentes sistemas de informação como a dificuldade mais restritiva ao trabalho.

Destaca-se a revisão do protocolo de terapia com prótons para câncer de próstata, o que deu origem a um modelo de informações clínicas e um fluxo de dados específico para o processo em questão.

Carvalho *et al* (2012) demonstram o potencial das técnicas de Mineração de Dados aplicando algoritmos de Classificação, descoberta de regras de associação e de agrupamento em uma base de dados clínicos sobre pacientes de fisioterapia a partir de dados sobre perfil profissional e informações clínicas. Os autores propõem que a decisão sobre a terapia aplicada seja apoiada pelos padrões encontrados, uma vez que o perfil da atividade profissional pode inferir os tipos de lesões possíveis, indicando as recomendações de prevenção ou de tratamento.

O artigo destaca as dificuldades relacionadas à qualidade das informações, sejam pela diferença de linguagem para preenchimento de campos textuais abertos, dados incompletos ou informações contraditórias. No contexto estudado pelas autoras, fica evidente a ausência de um processo apoiado por sistemas de informação para o trabalho do especialista em Saúde, o fisioterapeuta neste caso, o que resulta na perda das informações sobre pacientes e sobre a evolução do próprio tratamento.

Matínez e Bermúdez (2012) utilizam Redes Bayesianas e Árvores de Decisão para previsão e diagnóstico médico em doenças cardiovasculares com bases de dados cujos atributos são discretos.

Os autores defendem a combinação de diferentes algoritmos para a aplicação de Mineração de Dados no apoio ao especialista médico. No estudo apresentado, apresentam um aplicativo construído especificamente e defendem a simplicidade de navegação nas regras encontradas para o especialista em Saúde, uma vez que este não tem condições de dedicar muito tempo à análise dos padrões.

Concluem que o auxílio ao planejamento de tratamentos pode ser apoiado com Mineração de Dados desde que mediante a representação visual dos padrões de maneira personalizada e simplificada.

Moghimi *et al* (2012) constroem um sistema de apoio à decisão baseado na análise de riscos e previsão de resultados em cirurgias para implante de quadril e joelho, tendo como base o KDD, levando em conta que a decisão em realizar tais procedimentos deve considerar várias dimensões e os riscos sobre a qualidade de vida dos pacientes, a inovação das técnicas a ser empregada no implante, as condições de saúde do paciente, especialmente dos ossos a serem operados e os custos relativos ao tratamento em questão.

Para tanto, os autores adotam o fluxo de trabalho em cirurgias ortopédicas e a estratégia de encontrar as relações entre os fatores de riscos em si, mas também entre os riscos cirúrgicos e os resultados dos procedimentos, implementando um sistema de detecção de riscos em tempo real, baseado em Redes Neurais e Regras de Associação.

Os dados relativos aos resultados após as cirurgias são adicionados ao banco de dados e os padrões são atualizados por conta da atualização da base de dados.

Cruz-Ramírez *et al* (2013) aplicam algoritmos evolucionários com multiobjetivos e Redes Neurais para a determinação das melhores combinações doador-recipiente e para a previsão do sucesso e sobrevida em pacientes como suporte à decisão em procedimentos de transplante de fígado.

Os estudos foram conduzidos em bases de dados relativas a 11 unidades de transplante de fígado espanholas e recipientes com idade acima de 18 anos que receberam transplantes de doadores mortos, cujos dados foram obtidos durante o processo de transplante.

Os autores adotam um modelo de informações que represente os dados tanto para avaliação de compatibilidade entre doador e recipiente, quanto para a avaliação do estado de saúde do doador, o que permitiu o acompanhamento dos resultados dos transplantes até o decorrer de um ano após o transplante, a perda do enxerto por rejeição ou a morte do recipiente.

Os pontos de atenção indicados em cada publicação estão organizados no Quadro 1.

Quadro 1 Pontos de atenção para o uso efetivo de resultados da Mineração de Dados conforme as publicações analisadas

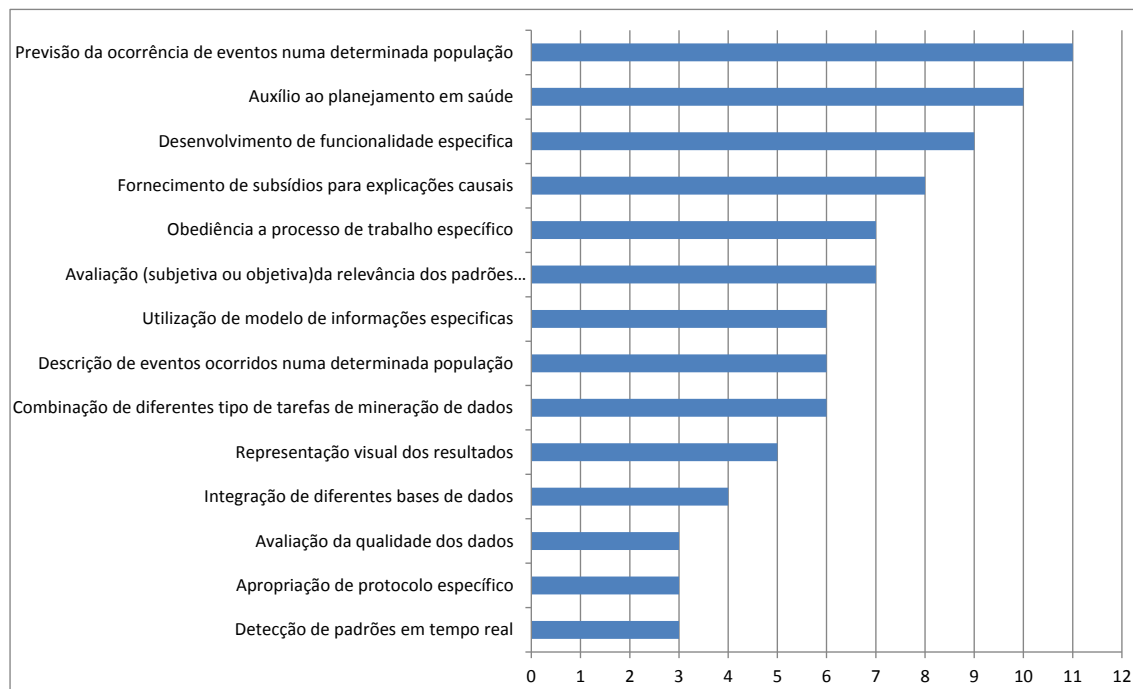
Pontos de atenção	West et al (2005)	Helma e Kasius (2006)	Kobus (2006)	Steiner et al. (2006)	Ramon et al (2007)	Stein Junior (2008)	Bodini Junior (2009)	Dallagassa (2009)	Kuretski (2009)	Meyfroidt (2009)	Ayed et al (2010)	Vianna et al (2010)	Lee at al (2011)	Carvalho et al (2012)	Martínes e Bermúdez (2012)	Moghimi et al (2012)	Trindade (2012)	Cruz-Ramírez (2013)
Apropriação de protocolo específico		X			X			X		X		X	X					
Avaliação da qualidade dos dados		X	X		X					X		X		X			X	
Integração de diferentes bases de dados					X					X		X		X				
Desenvolvimento de funcionalidade específica							X		X		X		X		X	X		
Obediência a processo de trabalho específico				X									X			X		
Utilização de modelo de informações específicas		X			X	X							X					X
Avaliação (subjetiva ou objetiva) da relevância dos padrões encontrados			X	X				X										

Combinação de diferentes tipo de tarefas de mineração de dados	X						X		X					X	X	X	X	
Deteção de padrões em tempo real					X					X	X							
Descrição de eventos ocorridos numa determinada população			X	X			X	X	X		X	X	X	X			X	
Previsão da ocorrência de eventos numa determinada população	X	X			X	X		X		X	X			X	X	X		X
Fornecimento de subsídios para explicações causais		X	X		X	X		X		X	X				X	X		
Auxílio ao planeamento em saúde			X		X			X		X					X	X	X	X
Representação visual dos resultados	X							X		X	X				X		X	

97

A partir da figura 1 é possível observar a frequência absoluta da presença dos pontos de atenção nas publicações analisadas.

Figura 1 Frequência dos pontos de atenção na bibliografia analisada



MINERAÇÃO SOBRE OS PONTOS DE ATENÇÃO IDENTIFICADOS

Extração de Regras de Associação. Os dados foram tratados com o algoritmo *Apriori* (AGRAWAL e SRIKANT, 1994), e foram obtidas 39114 regras de associação.

Pós-processamento. O conjunto de regras de associação foi tratado com o algoritmo DRE (MILANI E CARVALHO, 2013), obtendo-se 345 regras gerais acompanhadas de suas respectivas regras de exceção.

Dentre os pares de regras gerais e respectivas exceções encontradas, é possível destacar que o ponto de atenção Deteccção de Padrões em Tempo Real não relacionado com o Auxilio ao Planejamento em Saúde:

Se Auxilioplanejamentosaude = SIM *então* Deteccaotemporeal = NÃO
(Suporte = 44,4%; Confiança = 75%)

Entretanto, a regra acima possui 88 exceções, identificando que a deteccção de padrões em tempo real está presente nos relatos que incluíram integração de diferentes bases, avaliação da qualidade dos dados, apropriação de protocolo específico, explicações causais e previsão de eventos, como demonstrado no Quadro e 3.

Quadro 2 Frequência dos pontos de atenção associados à detecção de padrões em tempo real

Pontos de Atenção	Frequência nas regras de exceção
Integração de diferentes bases = SIM	52%
Avaliação da qualidade dos dados = SIM	36%
Apropriação de protocolo específico = SIM	34%
Explicações Causais = SIM	24%
Previsão de Eventos = SIM	22%

Quadro 3 Exemplos das Regras de exceção envolvendo detecção de padrões em tempo real

SE Explicações causais = SIM e Auxílio ao planejamento em saúde = SIM ENTÃO Detecção de padrões em tempo real = SIM (Suporte 11,1%; Confiança 100,0%)
SE Avaliação da qualidade dos dados = SIM e Apropriação de protocolo específico = SIM e Auxílio ao planejamento em saúde = SIM ENTÃO Detecção de padrões em tempo real = SIM (Suporte 11,1%; Confiança 100,0%)

Muito embora a Representação Visual dos Resultados não seja citada por 12 dos 18 estudos analisados, as regras de exceção obtidas demonstram que está presente quando explicações causais é um ponto de atenção citado pelos autores, estando presente em 45% das regras de exceção encontradas. (Quadro 4 e 5)

Quadro 4 Frequência dos pontos de atenção que levam à presença da representação visual dos resultados

Pontos de atenção	Frequência nas regras de exceção
Explicações causais SIM	45%
Previsão de eventos SIM	27%
Diferentes tarefas de DM SIM	24%
Apropriação de protocolo SIM	18%
Detecção em tempo real SIM	17%
Descrição de eventos ocorridos SIM	5%
Auxílio ao planejamento em saúde SIM	2%
Avaliação da qualidade dos dados SIM	1%

Quadro 5 Exemplos das Regras de exceção envolvendo detecção de Representação visual dos resultados

SE Modelo de informações específicas = NÃO e Diferentes tarefas de DM = NÃO e Explicações causais = SIM e Previsão de eventos = SIM ENTÃO Representação visual dos resultados = SIM (Suporte 16,7%; Confiança 100,0%)
SE Modelo de informações específicas = NÃO e Diferentes tarefas de DM=NÃO e Explicações causais=SIM ENTÃO Representação visual dos resultados=SIM (Suporte 22,2%; Confiança 75,0%)

5.DISCUSSÃO

A predominância de publicações encontradas (87%), que relatam a adoção de Mineração de Dados para a área clínica, evidencia um espaço importante da utilização desta técnica na rotina clínica para potencializar a eficiência do tratamento. Obstante a baixa frequência (23%) de relatos usando a Mineração de Dados na gestão da Saúde, esta pode proporcionar benefícios sobre critérios para a aplicação de recursos e elaboração de estratégias para melhor entender a população.

Por outro lado, o entendimento dos pontos de atenção para o uso da Mineração de Dados na Saúde permite a elaboração de táticas para aumento da eficiência e maior aproveitamento dos benefícios gerados. A necessidade por cuidados criteriosos na elaboração e execução de projetos de Mineração de Dados na Saúde fica evidente à medida que os materiais pesquisados demonstram que, assim como os padrões descobertos, a forma com que tais padrões são comprovados ou disponibilizados para os especialistas influencia sua aceitação e uso rotineiro. A existência de elementos específicos da Área da Saúde requer que requisitos de conformidade, qualidade ou de usabilidade sejam atendidos de maneira a aumentar a possibilidade de adoção intensiva dos resultados da Mineração de Dados.

Ainda que a descrição de padrões e a previsão de eventos sejam a natureza da Mineração de Dados, o fato de 17 estudos relatarem ambos como pontos de atenção mostra que o apoio à decisão em saúde requer aplicações que combinem capacidades descritivas e preditivas para a devida colaboração para o especialista, enfatizando possíveis causas, relações e consequências.

A prevalência dos objetivos primários da Mineração de Dados cercados por pontos relativos à representação visual dos resultados ou à detecção de padrões em tempo real deixa evidente que os algoritmos disponíveis ainda não atendem plenamente aos critérios de aceitação da área da Saúde. Um exemplo claro desta limitação está na necessidade de representar os resultados visualmente, mostrando que a saída dos algoritmos não é adequada para o contexto e exige tratamento posterior. Por outro lado, a existência de pontos de atenção relativos à integração de diferentes bases ou a avaliação da qualidade dos dados, também sugere que as entradas dos algoritmos atuais não são

suficientemente implementadas, requerendo um esforço na seleção, avaliação da qualidade e preparação dos dados.

Tais limitações impõem ao processo de Mineração de Dados maior custo operacional, pela necessidade de mais trabalhos e mais tempo, tanto para a preparação dos dados – pré-processamento – quanto para a representação dos resultados – pós-processamento e, ao mesmo tempo, abrem oportunidades para a construção de algoritmos ou até mesmo ferramentas para pré e pós-processamento, ampliando as funcionalidades e otimizando o processo, reservando tempo e espaço para a análise e interpretação dos padrões, contribuindo para a tomada de decisão mais eficiente.

O ponto de atenção Auxílio ao Planejamento em Saúde, por sua vez, mostra que a Mineração de Dados pode contribuir o ganho de eficiência e eficácia na decisão a médio e longo prazos, baseada em grandes volumes de informação.

O ponto Desenvolvimento de Funcionalidade Específica indica uma forte tendência a implantação de Sistemas de apoio à decisão elaborados especificamente e de forma amigável para o especialista em Saúde, aproximando-o dos possíveis benefícios da Mineração de Dados. Mas, esta possibilidade requer do pesquisador e desenvolvedor da solução atenção aos Protocolos Específicos, que orientam os procedimentos clínicos e ainda validam os resultados obtidos na própria Mineração, representando um desafio no entendimento mais profundo da área de aplicação.

A apropriação de protocolos específicos também indica a necessidade da área da Saúde em seguir princípios e procedimentos amplamente aceitos em cada especificidade médica. Este ponto de atenção, além de servir com base para testes e validações dos resultados, traz o desavio para o pesquisador e para o especialista de não permitirem que o conhecimento corrente determine os resultados, impedindo, assim, o surgimento de conhecimentos surpreendentes, mas sim, facilitando a validação dos modelos, sejam descritivos ou preditivos. Ou seja, o cuidado em não viciar os padrões obtidos com a Mineração de Dados amplia a complexidade da aplicação na área.

Muito embora os pontos de atenção Detecção de Padrões em Tempo Real e Apropriação de Protocolo Específico tenham as menores frequências nas publicações analisadas, o que evidencia afastamento da rotina médica, sua presença mostra que, além de importantes para a Mineração de Dados na Saúde, existem estudos para aproximar o KDD do dia a dia da Área da Saúde, indicando ricas possibilidades para exploração.

A associação entre os pontos de atenção identificados serve como base para estratégias que ampliem os benefícios da Mineração de Dados na Saúde mediante sua combinação. Por exemplo, a forte associação entre Detecção de Padrões em Tempo Real, Integração de Diferentes Bases de Dados e Avaliação da Qualidade dos Dados permite antecipar quais tarefas serão realizadas durante o processo de descoberta de conhecimentos e quais funcionalidades devem ser implementadas para acurácia e maior aceitação e uso na Área da Saúde.

Partindo do pressuposto de que a decisão pela combinação de pontos de atenção pode ser apoiada pela associação levantada neste estudo, é

possível induzir uma série de alternativas, compondo diferentes caminhos para o aumento da efetividade da Mineração de Dados na Saúde.

6.CONCLUSÃO

Utilizar a Mineração de Dados para analisar os grandes volumes de informações clínicas disponíveis pode auxiliar os especialistas em Saúde na melhora das condições dos pacientes e no ganho de eficiência no uso dos recursos médicos. Entretanto, mesmo sendo amplamente aplicada a problemas médicos, há uma necessidade evidente em lançar mão de estratégias e boas práticas que aumentem as chances da Mineração de Dados ser adotada na rotina médica.

Ainda que a adoção de Sistemas de Informação na Saúde seja evidente, a avalanche de informações clínicas geradas requer apoio computacional para a tomada de decisões. Há fortes evidências na literatura sobre a eficiência da Mineração de Dados na descoberta de conhecimentos, proporcionando eficiência, melhores tratamentos e ganho de qualidade de vida para os pacientes.

O planejamento efetivo de ações de prevenção ou de tratamento, mediante modelos que permitam ao especialista avaliar alternativas previamente (e automaticamente) elaboradas e indicar resultados futuros com alta acurácia, com explicações causais em tempo real, mais complexas e, por isso, mais completas, agiliza a resposta clínica, aumenta as chances de cura e reduz riscos para os pacientes. Embora seja possível identificar pontos de atenção para o uso da Mineração de Dados na Saúde, a aplicação de tais questões está ligada ao contexto e deve ser detalhadamente planejada e avaliada.

Por outro lado, o uso de Regras de Exceção para obter a associação entre os pontos identificados permite traçar estratégias mais complexas, atendendo um maior número de requisitos, para o uso rotineiro da Mineração de Dados.

Embora haja inúmeras experiências e relatos que indiquem quais podem ser os pontos de atenção para o sucesso no uso da Mineração de Dados na Saúde, fica evidente a ausência de um guia ou conjunto de padrões práticos para tal fim, o que denota uma série de oportunidades ricas e interessantes para a pesquisa sobre a aceitação de aplicação da Mineração de Dados em sistemas de apoio à decisão médica.

Referências

- AGRAWAL Rakesh; SRIKANT Ramakrishnan. **Fast Algorithms for Mining Association Rules in Large Databases**. 20th International Conference on Very Large Data Bases, 1994. p 478 – 499.
- AYED, Mounir Ben; Ltifi, Hela; Kolski, Christophe; Alimi, Adel M. **A user-centered approach for the design and implementation of KDD-based DSS: A case study in the healthcare domain**. Decision Support Systems, 2010. n50 p 64–78.
- BLONBERG, Luciano Costa. **Gestão de Métricas e Indicadores de Doenças em Saúde Bucal Suportado por um Ambiente de Descoberta de**

- Conhecimento em Banco de Dados.** Pontifícia Universidade Católica do Rio Grande do Sul, 2010. Dissertação de Mestrado. 107f.
- BODINI JUNIOR, Antonio Carlos. **Utilização de técnicas de data mining na detecção de outliers em auxílio à auditoria operacional com um estudo de caso com dados do Sistema de Informações Hospitalares.** Universidade Federal do Rio de Janeiro, 2009. Tese de Doutorado. 122 f.
- CARVALHO, Deborah; MOSER, Auristela Duarte; SILVA, Verônica Andrade; DALLAGASSA, Marcelo Rossano. **Mineração de Dados Aplicada à Fisioterapia.** Fisioterapia e Movimento, 2012. V25, n.3, p 595 – 605, jul / set.
- CRUZ-RAMÍREZ, Manuel; HERVÁS-MARTÍNEZ, César; GUTIÉRREZ, Pedro Antonio; PÉREZ-ORTIZ, María; BRICEÑO, Javier; DE LA MATA, Manuel. **Memetic Pareto differential evolutionary neural network used to solve an unbalanced liver transplantation problem.** Soft Computing, 2013. V17 n2 p 275-284.
- DALLAGASSA, Marcelo Rosano. **Concepção de uma metodologia para identificação de beneficiários com indicativos de Diabetes Melitus tipo 2.** Pontifícia Universidade Católica do Paraná. Curitiba, 2009. Dissertação de Mestrado. 105 f.
- FAYYAD, Usama; PIATETSKI-SHAPIO, Gregory; SMYTH, Padhraic. **The KDD Process for Extracting Useful Knowledge from Volumes of Data.** Communications of the ACM, 1996. n11 v39 27 – 34..
- HELMA, Christoph; KAZIUS, Jeroen. **Artificial Intelligence and Data Mining for Toxicity Prediction.** Current Computer Aided Drug Design, 2005. n2. v2. 123-133.
- HUSSAIN, Farhad; LIU, Huan; SUZUKI, Einoshin; LU, Hongjun. **Exception Rule Mining with Relative Interestingness Measure.** PAKDD, 2000. V1805 (1), 86 – 97.
- KOBUS, Luciana Schleder Gonçalves. **Aplicação da Descoberta de Conhecimento para identificação de usuários com doenças cardiovasculares elegíveis para programas de gerenciamento de caso.** Pontifícia Universidade Católica do Paraná. Curitiba, 2006. Dissertação de Mestrado 145 f.
- KURETZKI, Carlos Henrique.
- Técnicas de Mineração de Dados Aplicadas em Bases de Dados da Saúde a partir de Protocolos Eletrônicos.** Universidade Federal do Paraná. Curitiba, 2009. Dissertação de Mestrado 98 f.
- MARTÍNEZ, Guillermo Roberto Salarte; BERMÚDEZ, Yanci Viviana Castro. **Modelo híbrido para el diagnóstico de enfermedades cardiovasculares basado en inteligencia artificial.** Tecnura, 2012. v16 n33 p35 a 53.
- STEIN JUNIOR, Altair Von. **Descoberta de Regras por meio de KDD para a Classificação de Micro Áreas Homogêneas de Risco.** Pontifícia Universidade Católica do Paraná, Curitiba, 2008. Dissertação de Mestrado 106f.
- STEINER, Maria Terezinha Arns; SOMA, Ney Yoshihiro; SHIMIZU, Tamio; NIEVOLA, Júlio César. STEINER NETO, Pedro José. **Abordagem de um problema médico por meio do uso do processo KDD com ênfase em análise exploratória dos dados.** Revista Gestão e Produção, 2006. V13, n2, p325-337.
- RAMON Jan, FIERENS Daan; GÜIZA Fabián; MEYFROIDT Geert; BLOCKEEL Hendrik; BRUYNOOGHE Maurice; VAN DEN BERGHE, Greet. **Mining data**

from intensive care patients. Advanced Engineering Informatics, 2007. n21 p243–256.

TRINDADE, Carla Machado da. MORO, Claudia M. Cabral. ALDENUCCI, Márcia Gil. NIEVOLA, Júlio Cesar. CARVALHO, Deborah Ribeiro. **Technology in Health: Knowledge Discovery in Public Health Databases: Study of Viral Hepatitis in The State Of Paraná, Brazil.** Iberoamerican Journal of Applied Computing. 2012. n2 v2..

LE Anh H. LIU, Brent. SCHULTE Reinhard HUANG H. K. **Intelligent ePR system for evidence-based research in radiotherapy: proton therapy for prostate cancer.** International Journal of Computer Assisted Radiology and Surgery, 2011. V6, Issue 6 , p 769-784.

MARISCAL, Gonzalo; MARBÁN, Óscar; FERNANDEZ, Covadonga. **A survey of data mining and knowledge discovery process models and methodologies.** The Knowledge Engineering Review, 2010. V25:2, 137–166.

MEYFROIDT Geert; GÜIZA, Fabian; RAMON, Jan; BRUYNOOGHE, Maurice. **Machine learning techniques to examine large patient databases.** Best Practice & Research Clinical Anaesthesiology, 2009. N 23 p 127–143.

MILANI, Cristian. Simioni; CARVALHO, Deborah Ribeiro. **Pós-Processamento em KDD.** Revista de Engenharia e Tecnologia, , 2013. V 5, p. 151-162.

MOGHIMI, Hoda; ZADEH, Hossein; SCHAFFER, Jonathan; WICKRAMASINGHE, Nilmini. **Incorporating intelligent risk detection to enable superior decision support: the example of orthopaedic surgeries.** Heath and Techonology, 2012. V2, n1 p33-41.

QUINLAN, John Ross. **Introduction to Decision Trees.** Machine Learning, 1986. N1 p 81-106.

QUINLAN, John Ross. **C4.5: Programs for Machine Learning.** Morgan Kaufmann Publishers. 1993

VIANNA, Rossana .Cristina .Xavier .Ferreira.; MORO, Claudia Maria Cabral de Barra; MOYSES, Samuel Jorge.; CARVALHO, Deborah.; NIEVOLA, Júlio Cesar. **Mineração de Dados e características da mortalidade infantil.** Caderno Saúde Pública, Rio de Janeiro, 2010

WEST, David; MANGIAMELI, Paul; RAMPAL, Rohit; WEST, Vivian. **Ensemble strategies for a medical diagnostic decision support system: A breast cancer diagnosis application.** European Journal of Operational Research, 2005. V162.p 532–551.

WITtEN, Ian; FRANK, Eibe; HALL, Mark. **Data Mining: Practical Machine Learning Tools and Techniques.** Morgan Kaufmann. San Francisco. 2011.

APÊNDICE B

ASSOCIATION EXPLORER

Protótipo para exploração de regras de associação, padrões de causa e efeito e janelas de tempo entre procedimentos de saúde.

GUIA RÁPIDO PARA USO**1. INSTALAR O ASSOCIATION EXPLORER**

O Association Explorer não requer instalação, basta descompactar o arquivo ASSOCIATIONEXPLORER.RAR EM uma pasta no disco do computador.

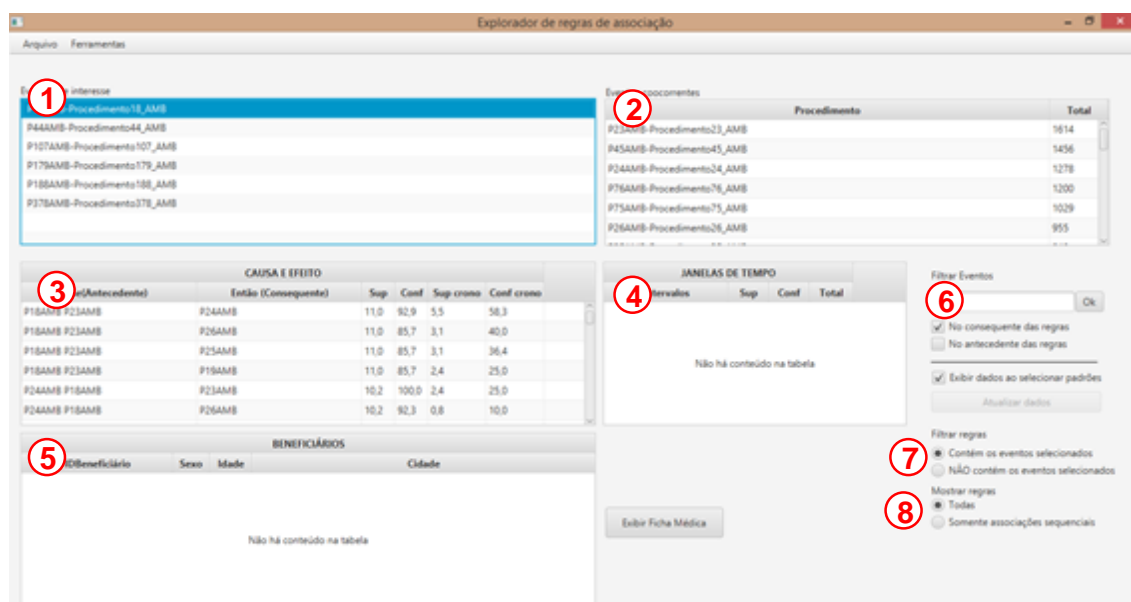
Os seguintes arquivos serão descompactados no disco do computador:

- AssociationExplorer.jar = arquivo executável do protótipo;
- AssocExpAmb.db = arquivo de dados de configuração do protótipo
- Minex.db = arquivo da base de conhecimento do protótipo
- Apriori.exe = executável para a obtenção de regras de associação
- Chrono_assoc.exe = executável para obtenção das associações cronológicas
- Interesse.txt =arquivo com códigos de procedimentos de interesse (para testes)
- Dados_PUC(2015) = pasta com arquivos de entrada com procedimentos para geração da base de conhecimento do protótipo (para testes)

2. INTERFACE DE NAVEGAÇÃO NOS PADRÕES

Janela principal para interação do modelo com o usuário, bem como para apresentação dos padrões e dados minerados.

Figura B.1 Interface de usuário do modelo



Fonte: O autor (2015)

A interação do usuário e a navegação nos padrões e nos dados dos beneficiários se darão mediante a manipulação dos seguintes componentes:

1: Lista de Eventos: exibe os eventos encontrados na base de dados e partir da qual o usuário deve selecionar o evento a ser analisado.

2: Lista de eventos coocorrentes. Esta lista é atualizada após a seleção de um dos eventos na Lista de Eventos (1) e tem a finalidade de informar quais são os eventos encontrados na base de dados que ocorrem juntamente com o evento sendo analisado.

3: Painel de Causa e Efeito: Este painel exibe as regras de associação obtidas acompanhadas de suporte e confiança e de suporte e confiança cronológicos.

4: Janelas de Tempo: Lista com os padrões de janelas de tempo identificados para a regra selecionada no painel Causa e Efeito (3).

5: Beneficiários. Lista com os dados dos beneficiários que satisfazem à filtragem pelos eventos presentes na regra selecionada no painel de Causa e Efeito (3).

6: Filtrar eventos. O usuário pode indicar se deseja que as regras sejam filtradas a partir do evento selecionado no antecedente, no consequente ou em ambos. Também é possível digitar parte do código do evento como critério de filtragem para a lista Eventos de interesse (1).

7: Exibir dados automaticamente. Quando selecionada, gráfico, padrões e dados são atualizados automaticamente ao selecionar um dos eventos da Lista de Eventos (1).

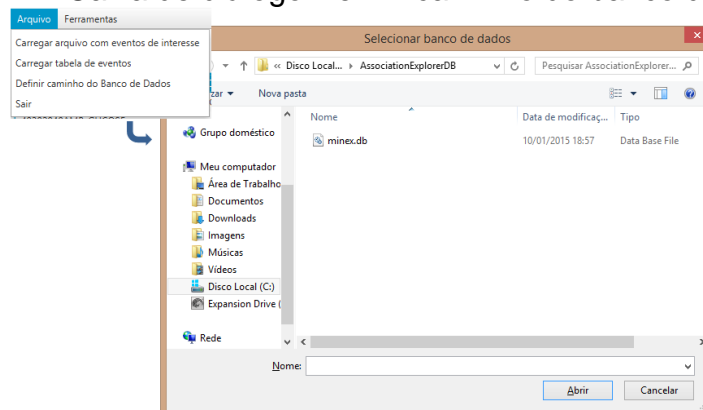
8: Filtrar regras: O usuário pode optar por mostrar as regras que possuem os eventos ou mostrar as regras que não possuem os eventos (conjunto exclusão).

3. INICIALIZAR O ASSOCIATION EXPLORER

Passo 1: Definir o caminho do banco de dados (MINEX.DB)

- Abra o AssociationExplorer.
- Menu Arquivo/Definir caminho do banco de dados;
- Selecione o arquivo MINEX.DB na pasta onde foi copiado e clique em Abrir

Figura B.2 Caixa de diálogo Definir caminho do banco de dados

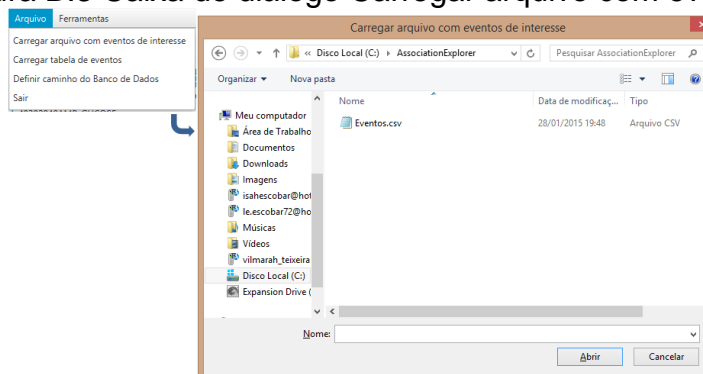


Fonte: O autor (2015)

Passo 2: Carregar o arquivo de eventos

- Menu Arquivo / Carregar tabela de eventos
- Selecione o arquivo com os eventos e clique em abrir.

Figura B.3 Caixa de diálogo Carregar arquivo com eventos



Fonte: O autor (2015)

O arquivo de eventos contém todas as transações sobre as quais se deseja obter padrões e realizar as análises de associação.

Nota: Deve-se garantir que os conteúdos dos campos do arquivo de eventos não possuam ponto-e-vírgula, apóstrofes ou ENTER.

Esses caracteres devem ser retirados dos dados antes de gerar o arquivo CSV.

Esta operação é demorada, em função do volume de dados e do processamento requerido para sua conversão!

Formato: Arquivo CSV (separado por ponto-e-vírgula)

Layout:

Quadro B.1 Layout do arquivo de entrada - Eventos

Item	Dado
1	CODIGO_USUARIO
2	SEXO
3	IDADE
4	CIDADE
5	DATA_ATENDIMENTO
6	CODIGO_SERVICO
7	TIPO_SERVICO
8	SERVICO
9	CODIGO_CID
10	NOME_PRESTADOR
11	LOCAL_ATENDIMENTO
12	QUANTIDADE
13	CUSTO

Fonte: O autor (2015)

4. OBTER OS PADRÕES A PARTIR DOS DADOS

- Menu Ferramentas/Obter Regras de Associação
- A tela de parametrização para obtenção dos padrões será aberta.

Figura B.4 Tela para parametrizar a obtenção dos padrões de associação

Obter Regras de Associação

Suporte mínimo:

Confiança mínima:

Mínimo Máximo

Itens na regra:

Eventos de Interesse: ...

Janelas de tempo:

Cancelar Obter regras

Fonte: O autor (2015)

- Deve-se informar:
 - Suporte mínimo para as Regras de Associação (Apriori)
 - Confiança Mínima para as Regras de Associação (Apriori)
 - Caminho para o arquivo com os códigos dos eventos de interesse, para obtenção das Janelas de Tempo
 - Intervalos, em dias, separados por ponto-e-vírgula, para as Janelas de Tempo
- Clicar no botão **Obter regras**

Atenção: Todos os padrões serão apagados da base de padrões e um novo conjunto será gravado.

Esta operação é demorada, em função do volume de dados e do processamento requerido para sua conversão!

Atenção: Se desejar alterar os eventos de interesse sem apagar a base de padrões, utilize:

Menu Arquivo / Carregar arquivo com eventos de interesse
(Texto simples com os códigos dos procedimentos separados por ENTER)

5. EXPLORAR OS PADRÕES OBTIDOS

Figura B.5 Janela de navegação do especialista

Eventos de interesse

P18AMB-Procedimento18_AMB
P44AMB-Procedimento44_AMB
P107AMB-Procedimento107_AMB
P179AMB-Procedimento179_AMB
P188AMB-Procedimento188_AMB
P378AMB-Procedimento378_AMB

Eventos coocorrentes

Procedimento	Total
P23AMB-Procedimento23_AMB	1614
P45AMB-Procedimento45_AMB	1456
P24AMB-Procedimento24_AMB	1278
P76AMB-Procedimento76_AMB	1200
P75AMB-Procedimento75_AMB	1029
P26AMB-Procedimento26_AMB	955

CAUSA E EFEITO

Se(Antecedente)	Então (Consequente)	Sup	Conf	Sup cromo	Conf cromo
P18AMB P23AMB	P24AMB	11,0	92,9	5,5	58,3
P18AMB P23AMB	P26AMB	11,0	85,7	3,1	40,0
P18AMB P23AMB	P25AMB	11,0	85,7	3,1	36,4
P18AMB P23AMB	P19AMB	11,0	85,7	2,4	25,0
P24AMB P18AMB	P23AMB	10,2	100,0	2,4	25,0
P24AMB P18AMB	P26AMB	10,2	92,3	0,8	10,0

JANELAS DE TEMPO

Intervalos	Sup	Conf	Total
Não há conteúdo na tabela			

BENEFICIÁRIOS

IDBeneficiário	Sexo	Idade	Cidade
Não há conteúdo na tabela			

Filtrar Eventos

☐ No consequente das regras
☐ No antecedente das regras
☒ Exibir dados ao selecionar padrões

Filtrar regras

☒ Contém os eventos selecionados
☐ NÃO contém os eventos selecionados

Mostrar regras

☒ Todas
☐ Somente associações sequenciais

Atualizar dados

Exibir Ficha Médica

Fonte: O autor (2015)

Passo 1 –Selecione um dos procedimentos na caixa EVENTOS DE INTERESSE

Figura B.6 Caixa de listagem Eventos de Interesse

Eventos de interesse

P18AMB-Procedimento18_AMB
P44AMB-Procedimento44_AMB
P107AMB-Procedimento107_AMB
P179AMB-Procedimento179_AMB
P188AMB-Procedimento188_AMB
P378AMB-Procedimento378_AMB

Fonte: O autor (2015)

- A contagem dos eventos associados será exibida na caixa **Eventos coocorrentes**.

Figura B.7 Caixa de listagem – Eventos coocorrentes

Procedimento	Total
P23AMB-Procedimento23_AMB	1614
P45AMB-Procedimento45_AMB	1456
P24AMB-Procedimento24_AMB	1278
P76AMB-Procedimento76_AMB	1200
P75AMB-Procedimento75_AMB	1029
P26AMB-Procedimento26_AMB	955

Fonte: O autor (2015)

- As Regras de Associação serão exibidas na caixa **CAUSA E EFEITO**

Passo 2 – Selecione uma das regras na caixa CAUSA E EFEITO

Figura B.8 Tabela Causa e Efeito – exibe as regras de associação e sequenciais

Se(Antecedente)	Então (Consequente)	Sup	Conf	Sup crono	Conf crono
P18AMB P23AMB	P24AMB	11,0	92,9	5,5	58,3
P18AMB P23AMB	P26AMB	11,0	85,7	3,1	40,0
P18AMB P23AMB	P25AMB	11,0	85,7	3,1	36,4
P18AMB P23AMB	P19AMB	11,0	85,7	2,4	25,0
P24AMB P18AMB	P23AMB	10,2	100,0	2,4	25,0
P24AMB P18AMB	P26AMB	10,2	92,3	0,8	10,0

Fonte: O autor (2015)

- Os padrões de Janela de Tempo serão exibidos na caixa **JANELAS DE TEMPO**

Figura B.9 Tabela Causa e Efeito – exibe as regras de associação e sequenciais

Intervalos	Sup	Conf	Total
Até 90 DIAS	8,7	100,0	11,0
Até 120 DIAS	8,7	100,0	11,0
Até 150 DIAS	8,7	100,0	11,0

Fonte: O autor (2015)

- Os dados dos beneficiários cobertos pela regra selecionada serão exibidos na caixa **BENEFICIÁRIOS**.

Passo 3 – Selecione um dos beneficiários na caixa BENEFICIÁRIOS e clique no botão EXIBIR FICHA MÉDICA

- O histórico de procedimentos liberados para o beneficiário selecionado será exibido em na janela Ficha Médica.

Figura B.10 Ficha médica do beneficiário

The screenshot shows a window titled 'Ficha médica'. At the top, there are input fields for 'Beneficiário' (containing 'Beneficiario 100'), 'Sexo' (containing 'M'), 'Idade' (containing '77 anos'), and 'Cidade' (containing 'PALOTINA'). Below these fields is a table with the following columns: 'Codigo serviço', 'Serviço', 'Local', 'Prestador', 'Data atend', 'Qtd', and 'Custo guia (R\$)'. The table contains 18 rows of data. At the bottom right of the window is a 'Voltar' button.

Codigo serviço	Serviço	Local	Prestador	Data atend	Qtd	Custo guia (R\$)
P73AMB	Procedimento73_AMB	Local 1026	PRESTADOR 114	2014-08-01	0.5	4.01
P19AMB	Procedimento19_AMB	Local 1026	PRESTADOR 114	2014-08-01	0.5	5.16
P633AMB	Procedimento633_AMB	Local 1026	PRESTADOR 114	2014-08-01	0.5	5.16
P25AMB	Procedimento25_AMB	Local 1026	PRESTADOR 114	2014-08-01	0.5	4.01
P26AMB	Procedimento26_AMB	Local 1026	PRESTADOR 114	2014-08-01	0.5	4.01
P23AMB	Procedimento23_AMB	Local 1026	PRESTADOR 114	2014-08-01	0.5	8.61
P179AMB	Procedimento179_AMB	Local 1026	PRESTADOR 114	2014-08-01	0.5	0.0
P19AMB	Procedimento19_AMB	Local 1026	PRESTADOR 114	2014-08-01	0.5	0.06
P633AMB	Procedimento633_AMB	Local 1026	PRESTADOR 114	2014-08-01	0.5	0.06
P25AMB	Procedimento25_AMB	Local 1026	PRESTADOR 114	2014-08-01	0.5	0.05
P26AMB	Procedimento26_AMB	Local 1026	PRESTADOR 114	2014-08-01	0.5	0.05
P23AMB	Procedimento23_AMB	Local 1026	PRESTADOR 114	2014-08-01	0.5	0.09
P179AMB	Procedimento179_AMB	Local 1026	PRESTADOR 114	2014-08-01	0.5	72.47
P73AMB	Procedimento73_AMB	Local 1026	PRESTADOR 114	2014-08-01	0.5	0.05

Fonte: O autor(2015)

6. FILTRAGEM E EXIBIÇÃO DOS PADRÕES

Figura B.10 Caixa de filtragemdo modelo proposto

The screenshot shows a dialog box titled 'Filtrar Eventos'. It contains several sections:

- A search bar with an 'Ok' button (labeled A).
- Two checked checkboxes: 'No conseqüente das regras' (labeled B) and 'No antecedente das regras'.
- A checked checkbox: 'Exibir dados ao selecionar padrões' (labeled C).
- An 'Atualizar dados' button.
- A section titled 'Filtrar regras' with two radio buttons: 'Contém os eventos selecionados' (selected, labeled D) and 'NÃO contém os eventos selecionados'.
- A section titled 'Mostrar regras' with two radio buttons: 'Todas' (selected, labeled E) and 'Somente implicações'.

Fonte: O autor (2105)

A – Digite parte do código do procedimento que deseja e clique em **Ok**. A lista **Eventos de Interesse** será filtrada de acordo com o critério informado.

Apague o critério e clique em **Ok** para remover o filtro da lista **Eventos de interesse**.

B – Defina se as regras exibidas devem conter o procedimento selecionado no consequente ou no antecedente (ou a combinação de ambos).

C – Ative ou desative a exibição automática dos dados após selecionar um procedimento na caixa Eventos de interesse ou uma regra na caixa CAUSA E EFEITO.

Se for desativado, utilize o botão **Atualizar dados** para mostrar os padrões e beneficiários cobertos pelo procedimento e pelas regras selecionadas.

D – Defina se as regras mostradas devem conter ou não conter o procedimento selecionado na caixa **Eventos de Interesse**.

E – Defina se as todas as regras devem ser mostradas ou se somente as regras que possuam suporte e confiança cronológicos maiores que 0 (zero) devem ser mostradas na caixa **CAUSA E EFEITO**.

APÊNDICE C

Instrumento para avaliação do modelo proposto

UNIVERSIDADE FEDERAL DO PARANÁ
Mestrado Interdisciplinar em Ciência, Gestão e Tecnologia da Informação
**AVALIAÇÃO DO MODELO PARA IDENTIFICAÇÃO DE PACIENTES DE
ALTO CUSTO**

Leandro Escobar

Professora Dra. Deborah Carvalho Ribeiro

Solicito a gentileza em participar de pesquisa sobre o uso do KDD para a identificação de pacientes de auto custo em operadoras de saúde.

Os resultados coletados e analisados por pesquisa lhe serão retornados oportunamente.

Antecipadamente agradeço a colaboração e fico à total disposição.

Leandro Escobar – 41 8423 5242 – l.escobar72@gmail.com

OS DADOS RELATIVOS AO AVALIADOR OU PARTICIPANTE DESTA PESQUISA SERÃO MANTIDOS SOB SIGILO.

Após a utilização do modelo disponibilizado como objeto desta avaliação, assinale o quanto concorda ou discorda com as afirmações abaixo:

Q1. O sistema mostra informações claramente e é fácil compreendê-las

- () Discordo
() Sou Indiferente
() Concordo

Se você discordou, justifique:

Q2. Aprendi rapidamente a utilizar a interface de navegação do sistema

- () Discordo
() Sou Indiferente
() Concordo

Se você discordou, justifique:

Q3. Utilizaria o sistema em minha rotina de trabalho

- () Discordo
() Sou Indiferente
() Concordo

Se você discordou, justifique:

Q4. A frequência dos eventos é útil para a análise dos padrões

- () Discordo
() Sou Indiferente
() Concordo

Se você discordou, justifique:

Q5. Os padrões apresentados são claros e facilmente compreendidos

- () Discordo
() Sou Indiferente
() Concordo

Se você discordou, justifique:

Q6. Os dados apresentados sobre os beneficiários são úteis e relevantes para meu trabalho

- () Discordo
() Sou Indiferente
() Concordo

Se você discordou, justifique:

Q7. A disposição dos padrões e informações é adequada para o uso em minha rotina

- () Discordo
() Sou Indiferente
() Concordo

Se você discordou, justifique:

Q8. Os padrões apresentados esclarecem como os eventos estão relacionados.

- () Discordo
() Sou Indiferente
() Concordo

Se você discordou, justifique:

Q9. As janelas de tempo permitem compreender como os eventos estão relacionados.

- () Discordo
() Sou Indiferente
() Concordo

Se você discordou, justifique:

Q10. Compreendi claramente os padrões apresentados e como eles descrevem os eventos ocorridos.

- () Discordo
() Sou Indiferente
() Concordo

Se você discordou, justifique:

Por favor, informe:

Q11. Sua área de formação/especialidade: _____

Q12. Há quanto tempo atua com auditoria? _____ anos.

Q13. Utiliza ferramentas estatísticas para a análise de procedimentos a serem auditados?

- () Sim () Não

Se respondeu **não**, descreva como realiza a análise de procedimentos.

APÊNDICE D

Quadros com as regras gerais e regras de exceção para o Ponto de Atenção Representação visual dos resultados, Descrição de eventos, Funcionalidade específica e Explicações causais.

1) Regras gerais e de exceção envolvendo a presença de Representação Visual dos Resultados

	Regra geral	Sup.	Conf.	Regras de exceção	Sup.	Conf.
1	Se Processo específico AUSENTE Então Representação visual dos resultados AUSENTE	83,3%	60,0%	Se Processo específico AUSENTE Detecção de padrões em tempo real PRESENTE Então Representação visual dos resultados PRESENTE	16,7%	66,7%
2	Se Avaliação da relevância dos padrões encontrados AUSENTE Então Representação visual dos resultados AUSENTE	83,3%	66,7%	Se Avaliação da relevância dos padrões encontrados AUSENTE e Detecção de padrões em tempo real PRESENTE Então Representação visual dos resultados PRESENTE	16,7%	66,7%
3	Se Integração diferentes bases AUSENTE Então Representação visual dos resultados AUSENTE	77,8%	64,3%	Se Integração diferentes bases AUSENTE e Detecção de padrões em tempo real PRESENTE Então Representação visual dos resultados PRESENTE	5,6%	100,0%
4	Se			Se		

	Apropriação de protocolo específico AUSENTE Então Representação visual dos resultados AUSENTE	66,7%	66,7%	Apropriação de protocolo específico AUSENTE e Detecção de padrões em tempo real PRESENTE Então Representação visual dos resultados PRESENTE	5,6%	100,0%
5	Se Avaliação da qualidade dos dados AUSENTE Então Representação visual dos resultados AUSENTE	61,10%	63,6%	Se Avaliação da qualidade dos dados AUSENTE e Detecção de padrões em tempo real PRESENTE Então Representação visual dos resultados PRESENTE	5,6%	100,0%
6	Se Combinação de Diferentes tarefas de Mineração de Dados AUSENTE Então Representação visual dos resultados AUSENTE	61,10%	72,7%	Se Combinação de Diferentes tarefas de Mineração de Dados AUSENTE e Detecção de padrões em tempo real PRESENTE Então Representação visual dos resultados PRESENTE	16,7%	66,7%
7	Se Auxílio ao planejamento em saúde AUSENTE Então Representação visual dos resultados AUSENTE	55,60%	80,0%	Se Auxílio ao planejamento em saúde AUSENTE e Detecção de padrões em tempo real PRESENTE Então Representação visual dos resultados PRESENTE	5,6%	100,0%
8	Se			Se		

	<p>Descrição de eventos ocorridos PRESENTE</p> <p>Então Representação visual dos resultados AUSENTE</p>	55,60%	70,0%	<p>Deteccção de padrões em tempo real PRESENTE e Descrição de eventos ocorridos PRESENTE</p> <p>Então Representação visual dos resultados PRESENTE</p>	5,6%	100,0%
9	<p>Se Desenvolvimento de funcionalidade específica PRESENTE</p> <p>Então Representação visual dos resultados AUSENTE</p>	33,30%	66,7%	<p>Se Deteccção de padrões em tempo real PRESENTE e Desenvolvimento de funcionalidade específica PRESENTE</p> <p>Então Representação visual dos resultados PRESENTE</p>	5,6%	100,0%
10	<p>Se Descrição de eventos ocorridos PRESENTE</p> <p>Então Representação visual dos resultados AUSENTE</p>	55,60%	70,0%	<p>Se Explicações causais PRESENTE e Descrição de eventos ocorridos PRESENTE</p> <p>Então Representação visual dos resultados PRESENTE</p>	16,7%	66,7%
				<p>Se Auxílio ao planeamento em saúde PRESENTE e Descrição de eventos ocorridos PRESENTE</p> <p>Então Representação visual dos resultados PRESENTE</p>	16,7%	66,7%
11	Se			Se		

Apropriação de protocolo específico PRESENTE	33,30%	66,7%	Apropriação de protocolo específico PRESENTE e Auxílio ao planejamento em saúde PRESENTE	16,7%	66,7%
Então Representação visual dos resultados AUSENTE			Então Representação visual dos resultados PRESENTE		

Fonte: O autor (2015)

2) Regras gerais e de exceção envolvendo a presença de “Descrição de Eventos Ocorridos”

	Regra geral	Sup.	Conf.	Regras de exceção	Sup.	Conf.
1	Se Previsão de Eventos PRESENTE Então Descrição de eventos ocorridos AUSENTE	55,60%	70,00 %	Se Previsão de Eventos PRESENTE e Avaliação da relevância dos padrões encontrados PRESENTE Então Descrição de eventos ocorridos PRESENTE	5,60%	100,00%
2	Se Explicações causais PRESENTE Então Descrição de eventos ocorridos AUSENTE	44,40%	62,50 %	Se Representacao visual dos resultados PRESENTE e Explicações causais PRESENTE Então Descrição de eventos ocorridos PRESENTE	16,70%	66,70%
				Se Explicações causais PRESENTE e Avaliação da relevância dos padrões encontrados	11,10%	100,00%

				PRESENTE		
				Então Descrição de eventos ocorridos PRESENTE		
3	Se Auxílio ao planejamento em saúde PRESENTE Então Descrição de eventos ocorridos AUSENTE	44,40%	62,50 %	Se Auxílio ao planejamento em saúde PRESENTE e Avaliação da relevância dos padrões encontrados PRESENTE Então Descrição de eventos ocorridos PRESENTE	11,10%	100,00%
4	Se Modelo de informações específicas PRESENTE Então Descrição de eventos ocorridos AUSENTE	27,80%	80,00 %	Se Funcionalidade específica PRESENTE e Modelo de informações específicas PRESENTE Então Descrição de eventos ocorridos PRESENTE	5,60%	100,00%
				Se Modelo de informações específicas PRESENTE e Processo específico PRESENTE Então Descrição de eventos ocorridos PRESENTE	5,60%	100,00%
5	Se Detecção de padrões em tempo real PRESENTE	16,70%	66,70 %	Se Detecção de padrões em tempo real PRESENTE e Funcionalidade específica PRESENTE	5,60%	100,00%

Então Descrição de eventos ocorridos AUSENTE			Então Descrição de eventos ocorridos PRESENTE		
---	--	--	---	--	--

3) Regras gerais e de exceção envolvendo a presença de Desenvolvimento de Funcionalidade Específica

	Regra geral	Sup.	Conf.	Regras de exceção	Sup.	Conf.
1	Se Avaliação da relevância dos padrões encontrados AUSENTE Então Funcionalidade específica AUSENTE	83,3%	60,00%	Se Avaliação da relevância dos padrões encontrados AUSENTE e Processo específico PRESENTE Então Funcionalidade específica PRESENTE	11,10%	100,00%
2	Se Detecção de padrões em tempo real AUSENTE Então Funcionalidade específica AUSENTE	83,3%	66,70%	Se Detecção de padrões em tempo real AUSENTE e Processo específico PRESENTE Então Funcionalidade específica PRESENTE	16,70%	66,70%
3	Se Representacaovisualresultados AUSENTE Então Funcionalidade específica AUSENTE	66,7%	66,70%	Se Representacaovisualresultados AUSENTE e Diferentes tarefas de Mineração de Dados PRESENTE Então Funcionalidade específica PRESENTE	22,20%	75,00%
				Se Representacaovisualresultados AUSENTE e Processo específico PRESENTE	16,70%	66,70%

				Então Funcionalidade específica PRESENTE		
4	Se Auxílio ao planejamento em saúde AUSENTE Então Funcionalidade específica AUSENTE	55,6%	60,00%	Se Auxílio ao planejamento em saúde AUSENTE e Detecção de padrões em tempo real PRESENTE Então Funcionalidade específica PRESENTE	5,60%	100,00%
5	Se Previsão de eventos PRESENTE Então Funcionalidade específica AUSENTE	55,6%	80,00%	Se Previsão de eventos PRESENTE e Processo específico PRESENTE Então Funcionalidade específica PRESENTE	5,60%	100,00%
6	Se Descrição de eventos ocorridos PRESENTE Então Funcionalidade específica AUSENTE	55,6%	60,00%	Se Descrição de eventos ocorridos PRESENTE e Modelo de informações específicas PRESENTE Então Funcionalidade específica PRESENTE	5,60%	100,00%
				Se Detecção de padrões em tempo real PRESENTE e Descrição de eventos ocorridos PRESENTE Então Funcionalidade específica PRESENTE	5,60%	100,00%
7	Se			Se		

	<p>Descrição de eventos ocorridos AUSENTE</p> <p>Então Funcionalidade específica AUSENTE</p>	44,4%	75,00%	<p>Descrição de eventos ocorridos AUSENTE e Diferentes tarefas de Mineração de Dados PRESENTE</p> <p>Então Funcionalidade específica PRESENTE</p>	16,70%	66,70%
				<p>Se Descrição de eventos ocorridos AUSENTE e Processo específico PRESENTE</p> <p>Então Funcionalidade específica PRESENTE</p>	5,60%	100,00%
8	<p>Se Explicações causais PRESENTE</p> <p>Então Funcionalidade específica AUSENTE</p>	44,4%	75,00%	<p>Se Explicações causais PRESENTE e Diferentes tarefas de Mineração de Dados PRESENTE</p> <p>Então Funcionalidade específica PRESENTE</p>	5,60%	100,00%
				<p>Se Explicações causais PRESENTE e Processo específico PRESENTE</p> <p>Então Funcionalidade específica PRESENTE</p>	5,60%	100,00%
9	<p>Se Auxílio ao planejamento em saúde PRESENTE</p> <p>Então Funcionalidade específica AUSENTE</p>	44,4%	75,00%	<p>Se Auxílio ao planejamento em saúde PRESENTE e Diferentes tarefas de Mineração de Dados PRESENTE</p> <p>Então Funcionalidade específica PRESENTE</p>	16,70%	66,70%
				Se		

				Auxílio ao planejamento em saúde PRESENTE e Processo específico PRESENTE Então Funcionalidade específica PRESENTE	5,60%	100,00%
10	Se Apropriação de protocolo específico PRESENTE Então Funcionalidade específica AUSENTE	33,3%	83,30%	Se Apropriação de protocolo específico PRESENTE e Processo específico PRESENTE Então Funcionalidade específica PRESENTE	5,60%	100,00%
	Se Modelo de informações específicas PRESENTE Então Funcionalidade específica AUSENTE	27,8%	80,00%	Se Descrição de eventos ocorridos PRESENTE e Modelo de informações específicas PRESENTE Então Funcionalidade específica PRESENTE	5,60%	100,00%
				Se Modelo de informações específicas PRESENTE e Processo específico PRESENTE Então Funcionalidade específica PRESENTE	5,60%	100,00%
11	Se Funcionalidade específica AUSENTE Detecção de padrões em tempo real PRESENTE Então Funcionalidade específica AUSENTE	16,7%	66,70%	Se Detecção de padrões em tempo real PRESENTE e Descrição de eventos ocorridos PRESENTE Então Funcionalidade específica PRESENTE	5,60%	100,00%

4) Regras gerais e de exceção envolvendo a presença de Explicações Causais

	Regra geral	Sup.	Conf.	Regras de exceção	Sup.	Conf.
1	Se Avaliação da relevância dos padrões encontrados AUSENTE Então Explicações causais AUSENTE	83,30%	60,00%	Se Avaliação da relevância dos padrões encontrados AUSENTE e Previsão de eventos PRESENTE Então Explicações causais PRESENTE	50,00%	66,70%
				Se Avaliação da relevância dos padrões encontrados AUSENTE e Apropriação de protocolo específico PRESENTE Então Explicações causais PRESENTE	27,80%	60,00%
				Se Avaliação da relevância dos padrões encontrados AUSENTE e Modelo de informações específicas PRESENTE Então Explicações causais PRESENTE	27,80%	60,00%
				Se Avaliação da relevância dos padrões encontrados AUSENTE e Detecção de padrões em tempo real PRESENTE Então Explicações causais PRESENTE	16,70%	100,00%
2	Se			Se		

	Explicações causais AUSENTE Detecção de padrões em tempo real AUSENTE Então	83,30%	66,70%	Detecção de padrões em tempo real AUSENTE e Avaliação da relevância dos padrões encontrados PRESENTE Então Explicações causais PRESENTE	16,70%	66,70%
3	Se Explicações causais AUSENTE Modelo de informações específicas AUSENTE	72,20%	61,50%	Se Modelo de informações específicas AUSENTE e Previsão de eventos PRESENTE Então Explicações causais PRESENTE	33,30%	66,70%
				Se Modelo de informações específicas AUSENTE e Auxílio ao planejamento em saúde PRESENTE Então Explicações causais PRESENTE	33,30%	66,70%
				Se Modelo de informações específicas AUSENTE e Apropriação de protocolo específico PRESENTE Então Explicações causais PRESENTE	16,70%	66,70%
				Se Modelo de informações específicas AUSENTE e Avaliação da relevância dos padrões encontrados PRESENTE Então	16,70%	66,70%

				Explicações causais PRESENTE		
				Se Modelo de informações específicas AUSENTE e Detecção de padrões em tempo real PRESENTE Então Explicações causais PRESENTE	11,10%	100,00%
4	Se Apropriação de protocolo específico AUSENTE Então Explicações causais AUSENTE Se Avaliação da qualidade dos dados AUSENTE Então Explicações causais AUSENTE	66,70%	66,70%	Se Apropriação de protocolo específico AUSENTE e Detecção de padrões em tempo real PRESENTE Então Explicações causais PRESENTE	5,60%	100,00%
		61,10%	63,60%	Se Avaliação da qualidade dos dados AUSENTE e Previsão de eventos PRESENTE Então Explicações causais PRESENTE	33,30%	66,70%
				Se Avaliação da qualidade dos dados AUSENTE e Detecção de padrões em tempo real PRESENTE Então Explicações causais PRESENTE	5,60%	100,00%
5	Se Auxílio ao planejamento em saúde AUSENTE Então	55,60%	70,00%	Se Auxílio ao planejamento em saúde AUSENTE e Previsão de eventos PRESENTE Então	27,80%	60,00%

	Explicações causais AUSENTE			Explicações causais PRESENTE		
				Se Auxílio ao planejamento em saúde AUSENTE e Modelo de informações específicas PRESENTE Então Explicações causais PRESENTE	16,70%	66,70%
				Se Auxílio ao planejamento em saúde AUSENTE e Detecção de padrões em tempo real PRESENTE Então Explicações causais PRESENTE	5,60%	100,00%
6	Se Descrição de eventos ocorridos PRESENTE Então Explicações causais AUSENTE	55,60%	70,00%	Se Descrição de eventos ocorridos PRESENTE e Previsão de eventos PRESENTE Então Explicações causais PRESENTE	16,70%	66,70%
				Se Auxílio ao planejamento em saúde PRESENTE e Descrição de eventos ocorridos PRESENTE Então Explicações causais PRESENTE	16,70%	66,70%
				Se Representação visual dos resultados PRESENTE e Descrição de eventos ocorridos PRESENTE Então Explicações causais PRESENTE	16,70%	66,70%

				Se Descrição de eventos ocorridos PRESENTE e Avaliação da relevância dos padrões encontrados PRESENTE Então Explicações causais PRESENTE	16,70%	66,70%
				Se Detecção de padrões em tempo real PRESENTE e Descrição de eventos ocorridos PRESENTE Então Explicações causais PRESENTE	5,60%	100,00%
7	Se Diferentes tarefas de Mineração de Dados PRESENTE Então Explicações causais AUSENTE	38,90%	85,70%	Se Diferentes tarefas de Mineração de Dados PRESENTE e Processo específico PRESENTE Então Explicações causais PRESENTE	5,60%	100,00%
8	Se Funcionalidade específica PRESENTE Então Explicações causais AUSENTE	33,30%	66,70%	Se Previsão de eventos PRESENTE e Funcionalidade específica PRESENTE Então Explicações causais PRESENTE	11,10%	100,00%
				Se Detecção tempo real PRESENTE e Funcionalidade específica PRESENTE Então Explicações causais PRESENTE	5,60%	100,00%
9	Se			Se		

	Processo específico PRESENTE	16,70%	66,70%	Previsão de eventos PRESENTE e Processo específico PRESENTE	5,60%	100,00%
	Então			Então		
	Explicações causais AUSENTE			Explicações causais PRESENTE		
				Se Auxílio ao planejamento em saúde PRESENTE e Processo específico PRESENTE	5,60%	100,00%
				Então Explicações causais PRESENTE		
				Se Diferentes tarefas de Mineração de Dados PRESENTE e Processo específico PRESENTE	5,60%	100,00%
				Então Explicações causais PRESENTE		

APÊNDICE E

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

Você está sendo convidado(a) como voluntário(a) a participar da pesquisa:
PÓS-PROCESSAMENTO DE PADRÕES PARA IDENTIFICAÇÃO DE BENEFICIÁRIOS DE ALTO CUSTO EM OPERADORAS DE SAÚDE.

JUSTIFICATIVA, OS OBJETIVOS E OS PROCEDIMENTOS: O motivo que nos levam a estudar este problema é a constatação de que o KDD é pouco utilizado na rotina da Gestão em Saúde e, por outro lado, o KDD pode contribuir para a análise dos padrões relativos a beneficiário de alto custo., a pesquisa se justifica porque o KDD pode descobrir padrões ocultos em bases de dados, permitindo que a identificação de casos de alto custo seja eficiente e eficaz, auxiliando na elaboração de estratégias de prevenção de doenças e na otimização dos recursos disponíveis. O objetivo desse projeto é desenvolver um modelo que permita a navegação em informações obtidas com o KDD para identificar beneficiários de alto custo em operadoras de Saúde.

O(os) procedimento(s) de coleta de dados serão da seguinte forma:

- a) você será orientado sobre como utilizar o sistema desenvolvido para exploração das informações;
- b) você deverá utilizar o sistema desenvolvido, de forma a explorar as informações disponibilizadas e
- c) você deverá responder a um questionário com 13 perguntas sobre sua opinião quanto ao modelo desenvolvido.

DESCONFORTOS E RISCOS E BENEFÍCIOS:

Existe um desconforto mínimo para você, por conta do tempo cedido para participar desta pesquisa, mas que se justifica pela vantagem de se identificar beneficiários de alto custo com mais eficiência.

FORMA DE ACOMPANHAMENTO E ASSINTÊNCIA:

Caso você tenha alguma dificuldade ou dúvida no uso do sistema desenvolvido, estarei à sua disposição, acompanhando todos os passos a serem realizados pessoalmente. Assim, suas questões serão resolvidas imediatamente.

GARANTIA DE ESCLARECIMENTO, LIBERDADE DE RECUSA E

GARANTIA DE SIGILO: Você será esclarecido(a) sobre a pesquisa em qualquer aspecto que desejar. Você é livre para recusar-se a participar, retirar seu consentimento ou interromper a participação a qualquer momento. A sua participação é voluntária e a recusa em participar não irá acarretar qualquer penalidade ou perda de benefícios.

O pesquisador irá tratar a sua identidade com padrões profissionais de sigilo. Os resultados da tabulação dos questionários serão enviados para você e suas respostas permanecerão confidenciais. Seu nome ou o material que indique a sua participação não será liberado sem a sua permissão. Você não será identificado(a) em nenhuma publicação que possa resultar deste estudo. Uma cópia deste consentimento informado será arquivada no Curso Mestrado em Tecnologia em Saúde da Pontifícia Universidade Católica do Paraná e no Curso Mestrado em Ciência, Gestão e Tecnologia da informação da Universidade Federal de São Paulo e outra será fornecida a você.

CUSTOS DA PARTICIPAÇÃO, RESSARCIMENTO E INDENIZAÇÃO POR

EVENTUAIS DANOS: A participação no estudo não acarretará custos para você e não será disponível nenhuma compensação financeira adicional.

DECLARAÇÃO DO PARTICIPANTE

Eu, _____ fui informada (o) dos objetivos da pesquisa acima de maneira clara e detalhada e esclareci minhas dúvidas. Sei que em qualquer momento poderei solicitar novas informações e motivar minha decisão se assim o desejar. A professora orientadora **Dr Deborah Ribeiro Carvalho** certificou-me de que todos os dados desta pesquisa serão confidenciais.

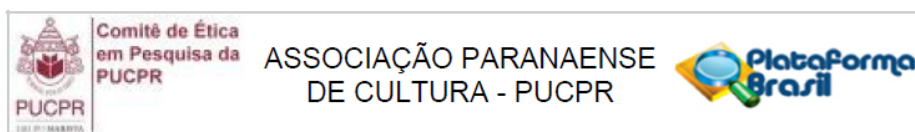
Também sei que caso existam gastos adicionais, estes serão absorvidos pelo orçamento da pesquisa. Em caso de dúvidas poderei chamar a estudante Leandro Fabian Almeida Escobar ou professora orientador no telefone (41) 8423 5242 ou o Comitê de Ética em Pesquisa da Faculdade de Saúde Pública da Universidade de São Paulo, sito à Av. Dr. Arnaldo, 715, Cerqueira César – São Paulo, SP.

Declaro que concordo em participar desse estudo. Recebi uma cópia deste termo de consentimento livre e esclarecido e me foi dada a oportunidade de ler e esclarecer as minhas dúvidas.

Nome	Assinatura do Participante	Data
Nome	Assinatura do Pesquisador	Data
Nome	Assinatura da Testemunha	Data

ANEXOS

Anexo 1 – Parecer consubstanciado do comitê de ética em pesquisa



PARECER CONSUBSTANCIADO DO CEP

DADOS DO PROJETO DE PESQUISA

Título da Pesquisa: PÓS-PROCESSAMENTO DE PADRÕES PARA IDENTIFICAÇÃO DE BENEFICIÁRIOS DE ALTO CUSTO EM OPERADORAS DE SAÚDE

Pesquisador: Deborah Ribeiro Carvalho

Área Temática:

Versão: 1

CAAE: 37962114.5.0000.0020

Instituição Proponente: Pontifícia Universidade Católica do Parana - PUCPR

Patrocinador Principal: Associação Paranaense de Cultura - PUCPR

DADOS DO PARECER

Número do Parecer: 864.236

Data da Relatoria: 04/11/2014

Apresentação do Projeto:

Esse trabalho apresenta um protótipo para integração de padrões minerados para a identificação de beneficiários de alto custo em operadora de saúde do Paraná, mediante o desenho de um modelo que permita a generalização de estratégias para fomento de uso de KDD na rotina em saúde.

Objetivo da Pesquisa:

Propor um modelo que integre padrões obtidos com o processo de descoberta de conhecimento -KDD- para identificação de beneficiários de alto custo em operadoras de Saúde.

Avaliação dos Riscos e Benefícios:

Riscos com relação à invasão de privacidade dos beneficiários e a contribuição está na organização dos pontos de atenção relevantes para a adoção do processo de descoberta de conhecimento em base de dados na rotina em Saúde.

Comentários e Considerações sobre a Pesquisa:

O projeto contempla as informações necessárias para sua análise, conforme o Comitê de Ética.

Considerações sobre os Termos de apresentação obrigatória:

O termo TCUD foi apresentado conforme exigência; foi apresentado TCLE, não necessário.

Endereço: Rua Imaculada Conceição 1155
Bairro: Prado Velho CEP: 80.215-901
UF: PR Município: CURITIBA
Telefone: (41)3271-2103 Fax: (41)3271-2103 E-mail: nep@pucpr.br

Continuação do Parecer: 864.236

Recomendações:

Este projeto foi aprovado conforme exigências do Comitê de Ética com algumas fortes recomendações: Retirar o TCLE pois não será necessário; o projeto foi aprovado somente para pesquisa em banco de dados. O TCLE apresentado está incorreto e não deverá ser usado. Recomenda-se revisão ortográfica no TCUD.

Conclusões ou Pendências e Lista de Inadequações:

O projeto foi aprovado e devem ser observadas as recomendações.

Situação do Parecer:

Aprovado

Necessita Apreciação da CONEP:

Não

Considerações Finais a critério do CEP:

CURITIBA, 10 de Novembro de 2014

Assinado por:
NAIM AKEL FILHO
(Coordenador)

Endereço: Rua Imaculada Conceição 1155		
Bairro: Prado Velho	CEP: 80.215-901	
UF: PR	Município: CURITIBA	
Telefone: (41)3271-2103	Fax: (41)3271-2103	E-mail: nep@pucpr.br